Islands and Integrals: Processes of Diversification in an Island Archipelago and Bayesian Methods of Comparative Phylogeographical Model Choice

By

Jamie R. Oaks

Submitted to the graduate degree program in Ecology and Evolutionary Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairperson Dr. Mark T. Holder

Chairperson Dr. Rafe M. Brown

Dr. John K. Kelly

Dr. Robert G. Moyle

Dr. Alan Redd

Date Defended: Oct 16, 2013

The Dissertation Committee for Jamie R. Oaks certifies that this is the approved version of the following dissertation:

Islands and Integrals: Processes of Diversification in an Island Archipelago and Bayesian Methods of Comparative Phylogeographical Model Choice

Chairperson Dr. Mark T. Holder

Chairperson Dr. Rafe M. Brown

Date Approved: Oct 16, 2013

Part I

Abstract

Abstract

Understanding the processes that generate, maintain, and regulate the assembly of biodiversity is a major goal of evolutionary biology. An important component of this goal is understanding how large-scale processes cause diversification across entire communities of species. These processes include geological and climatic mechanisms that alter the landscape and environment across which populations of organisms are distributed. Given the dynamic nature of our planet, such large-scale historical processes are likely common across most ecosystems, making them potential key drivers of diversification and community assembly. By simultaneously affecting entire communities of species, large-scale geological and climatic events are expected to generate patterns of divergences that are temporally clustered across affected groups of species. Such pulses of speciation are expected to leave a signature in the genetic variation within and among these lineages. This signal provides us with an opportunity to understand how past regional and global biogeographical processes have affected diversification by estimating the temporal patterns of divergence across present-day co-distributed species. The primary goal of this work is to (1) better understand processes of diversification within the Philippine Islands using DNA sequence data from a diverse set of vertebrate species, and (2) advance statistical methods of comparative phylogeographical model choice in order to improve the estimation of diversification models from genetic data.

The Philippine Islands are a particularly interesting system in which to address questions of diversification processes. Over the past few million years, the Islands of the Philippines have been repeatedly joined and fragmented due to oscillations in sea levels associated with

glacial cycles. It has been hypothesized that inter-glacial rises in sea level caused bouts of speciation across the islands due to the fragmentation of populations distributed across adjacent islands. In the first chapter of this dissertation, we test this hypothesis by applying a popular approximate-Bayesian method of phylogeographical model choice to infer the distribution of divergence times across a diverse set of 22-vertebrate taxa distributed across the Philippine Islands. Consistent with the sea-level driven model of diversification, the results strongly support recent and highly clustered divergences shared across the 22 population pairs. However, we also perform a suite of simulation-based analyses to assess the behavior of the method and find it to be biased toward supporting models with less parameter space and thus small numbers of divergence events shared across taxa.

In response to our findings in Chapter 1, a modification of the model-choice method was proposed as a means of circumventing the biases we reported. In Chapter 2, we used empirical and simulation-based analyses to investigate the behavior of this proposed method. We find the approach is still biased toward models with less parameter space, which can manifest in a strong tendency to sample predominantly from models that exclude the true values of the model's parameters. We also find that the bias toward small models still causes the method to prefer overly clustered models of divergence.

In Chapter 3, we introduce a new approximate-Bayesian model for comparative phylogeographical model-choice that estimates the temporal distribution of divergences across taxa from multi-locus DNA sequence data. By reparameterizing the model used in Chapter 1, and using more flexible priors on divergence models and nuisance parameters, we improve the robustness, accuracy, and power of the method for estimating the posterior probabilities of models of divergence across taxa. Our results demonstrate that the bias of the original model toward inferring models of clustered divergences is caused by a combination of (1) uniform priors on nuisance parameters reducing the marginal likelihoods of models with more divergence time parameters, and (2) a prior on divergence models that disfavors models with intermediate numbers of divergence time parameters.

In Chapter 4, we explore broad-scale temporal patterns of colonization and diversification of vertebrate groups in the Philippines. We mine the literature for estimates, and data to obtain estimates, of clade ages for a diverse set of vertebrate groups within the islands. We test whether time of colonization explains vertebrate species diversity in the islands (i.e., a time-for-diversification effect). Furthermore, we establish general patterns of colonization times and diversity across major vertebrate groups and discuss their implications in the formation of the archipelago's impressive biodiversity.

Part II

Acknowledgements

Acknowledgements

I thank the National Science Foundation (DEB 1011423), University of Kansas (KU) Office of Graduate Studies, Society of Systematic Biologists, Sigma Xi Scientific Research Society, KU Department of Ecology and Evolutionary Biology, and the KU Biodiversity Institute for financial support of this work. I also thank Mark Holder, the KU Information and Telecommunication Technology Center, KU Computing Center, and the iPlant Collaborative for the computational support necessary to conduct the analyses presented herein.

The patience and guidance of my committee members, John Kelly, Rob Moyle, and Alan Redd, greatly improved the quality of this dissertation. Rafe Brown and Mark Holder have been great advisors, colleagues, and friends. This work would not have been possible without their generosity, advice, and enthusiasm.

Over the past six years, I have been surrounded by a wonderful group of friends, including fellow students, postdocs, and faculty. My interactions with them over the years have contributed greatly to my intellectual growth, and have made my time in Lawrence, Kansas very enjoyable.

I thank my family for fostering my early interest in the natural world. Without their love and support, this work never would have been conducted. The incredible support and patience of my beautiful wife, Melissa, allowed me to complete this dissertation. She, and our dogs and cats, kept life fun and helped me maintain perspective and sanity when work was difficult.

Lastly, in the broadest sense, this dissertation is a product of the whole. Everyone has

played a role in allowing this work to happen and deserves full credit for the final product. I merely benefited from the luxury of doing what I enjoy.

Part III

Table of Contents

Contents

Part IV

Dissertation

Chapter 1

Evidence for climate-driven diversification? A caution for interpreting ABC inferences of simultaneous historical events

1.1 Introduction

Approximate Bayesian computation (ABC) is a statistical technique burgeoning in many subfields of biology due to its flexibility and ease of accommodating complex, parameter-rich models without the need of calculating a likelihood. (see Beaumont, 2010; Bertorelle et al., 2010; Csilléry et al., 2010, for reviews). The technique approximates the posterior of a model by accumulating samples of parameters from the prior that yield summary statistics similar to the values taken by these statistics on the observed data. The parameter estimates are often regression-adjusted to improve the approximation by accounting for variation in the probability of the data across the parameter space of the retained sample (Beaumont et al., 2002; Leuenberger and Wegmann, 2010; Blum and François, 2009).

One popular implementation of the ABC algorithm, `msBayes` (Huang et al., 2011), provides a statistical method for testing biogeographic hypotheses that predict temporally clustered divergences among co-distributed groups of organisms. Specifically, the `msBayes` model infers the distribution of divergence times among pairs of populations. Throughout this paper, we use "clustered", "simultaneous", and "co-divergence" interchangeably to describe the situation where `msBayes` infers the same time of divergence for any subset of population pairs.

In applications of `msBayes`, researchers have often found support for temporally clustered divergences among co-distributed pairs of taxa (Barber and Klicka, 2010; Bell et al., 2012; Carnaval et al., 2009; Chan et al., 2011; Daza et al., 2010; Hickerson et al., 2006; Huang et al., 2011; Lawson, 2010; Leaché et al., 2007; Plouviez et al., 2009; Stone et al., 2012; Voje et al., 2009). However, previous investigators have not performed power analyses to inform their interpretation of shared divergence times. Rather, support for co-divergence has been taken as support for a shared event, without determining how much variation in divergence times is permissible while still leading to an inference of "simultaneous" divergence. In this study, we use simulations based on an empirical dataset from the Philippines to determine the power of the ABC method implemented in `msBayes` for detecting temporal variation among divergences.

1.1.1 Pleistocene model of diversification in an Island Archipelago

The 7100+ islands of the Philippines may harbor the highest concentration of biodiversity on Earth (Brown and Diesmos, 2009; Heaney and Regalado, 1998), and have a relatively well-understood geologic history (Dickerson, 1928; Hall, 1998; Heaney, 1985; Inger, 1954; Voris, 2000; Yumul et al., 2008). During Pleistocene glacial cycles, sea-level fluctuations caused groups of previously isolated islands in the Philippines to undergo repeated cycles of connectivity and isolation (Voris, 2000). During glacial periods, when sea levels dropped to 120 meters below current levels, neighboring islands coalesced into seven main landmasses

known as Pleistocene Aggregate Island Complexes (PAICs; Brown and Diesmos, 2002). In interglacial periods, rising sea levels split the PAICs into the set of islands we see today. There have been at least six of these climate-driven cycles during the last 500,000 years (Rohling et al., 1998; Siddall et al., 2003), with additional cycles occurring in the late Pliocene and early Pleistocene (Haq et al., 1987; Miller et al., 2005).

The repeated formation and fragmentation of PAICs has been proposed as a mechanism of diversification across the Philippine Islands. (Heaney, 1986; Brown and Diesmos, 2002, 2009). The PAIC model makes a specific prediction: If repeated bouts of connectivity and isolation promoted diversification, divergence times between populations on islands connected during glacial lowstands should be clustered and correspond to when sea levels rose. If the PAIC cycles did not cause diversification, then divergences among island populations must be dispersal-mediated, and would not be temporally clustered across different groups. We test this prediction by inferring the temporal distribution of divergences among 22 population pairs from a diverse set of vertebrate taxa using mitochondrial sequence data from nearly one thousand individuals from across the Philippines.

1.2 Methods

1.2.1 Overview of the data

We gathered datasets for which there were mitochondrial sequence data for multiple individuals from populations of a species (or two closely related species) from two different islands within a PAIC. By maximizing population sample sizes and avoiding multiple use of data, we ended up with 470 individuals from 22 population pairs spanning five orders of terrestrial vertebrates. Thirteen pairs are from the Greater Mindanao PAIC and nine are from the Greater Negros-Panay PAIC (Table S5.3). We used these samples to infer the pattern of divergence times for the 22 population pairs with `msBayes`. We used 499 additional samples to estimate appropriate models of nucleotide substitution, gene trees, and the population

mutation rate (θ). These samples are from other island populations of the same species and closely related species (Table S5.8).

All megachiropteran sequences are from Roberts (2006a,b). All *Crocidura* shrew sequences are from Esselstyn et al. (2009). We included sequence data from geckos of the genera *Cyrtodactylus* and *Gekko* (Siler et al., 2010, 2012), frogs of the genus *Limnonectes* (Evans et al., 2003), bats of the genus *Hipposideros* (Esselstyn et al., 2012), and *Sphenomorphus*-group scincid lizards (currently assigned to the genera *Pinoyscincus* and *Insulasaurus*; Linkem et al., 2010, 2011), and augmented these datasets using the protocols in these papers to collect additional sequences. We collected sequences from snakes of the genus *Dendrelaphis* based on the protocols in Linkem et al. (2010) using the primers from Burbrink et al. (2000). All samples and their corresponding Genbank accession numbers are in Table S5.8.

1.2.2 Phylogenetic analyses

For phylogenetic analyses, we grouped closely related taxa, resulting in a total of 11 datasets that were easily aligned using MUSCLE (v3.7; Edgar, 2004) with no or few gaps (Dryad doi: 10.5061/dryad.5s07m). We used RAxML (v7.0.4; Stamatakis, 2006) to estimate maximum likelihood (ML) trees for each of the 11 alignments, using the rapid hill-climbing heuristic algorithm (Stamatakis et al., 2007). For each RAxML analysis, we ran 100 search replicates, applied the 'GTRMIX' model of nucleotide substitution, and used random starting trees. On each of the 11 ML trees inferred by RAxML, we estimated parameters of the HKY85 model (Hasegawa et al., 1985) using PAUP* (v4.0b10; Swofford, 2003). We used these HKY85 parameter estimates in subsequent `msBayes` analyses.

We used the Bayesian information criterion (BIC; Schwarz, 1978) to select the best-fit model of nucleotide substitution for each alignment using PAUP* and ModelTest (v3.7; Posada and Crandall, 1998). We inferred an ultrametric tree for each alignment in BEAST (v1.5.4; Drummond and Rambaut, 2007) using the BIC-selected model and a constant substitution rate of 2×10^{-8} per site per year. For each dataset, we ran two independent BEAST

analyses for 20 million generations and, after a conservative burn-in period of 5 million generations, sampled parameter values from the chain every 5,000 generations. We assessed stationarity by plotting the sampled parameter values and likelihood scores of both independent chains over generations, and confirming congruence between consensus trees from both posterior samples.

1.2.3 Estimates of θ to guide prior specification

Broad uniform priors may cause low marginal likelihoods for complex models, leading Bayesian model selection procedures to prefer overly simplistic models (Lindley, 1957; Jeffreys, 1961). We therefore estimated the population mutation rate ($\theta = N_e\mu$) using 499 samples from 30 populations (Table S5.8), and subsequently used these estimates to guide our choice of prior distributions of θ in the `msBayes` analyses. None of these 30 populations were used in the `msBayes` analyses, thus we avoided statistical problems associated with multiple use of data. All θ estimates were derived from the same gene fragment that was used for the corresponding taxon pair in the `msBayes` analyses. We used Dendropy (v3.2.1; Sukumaran and Holder, 2010) to calculate Watterson's θ (θ_W; Watterson, 1975) and the average per site nucleotide differences (π; Nei and Li, 1979).

1.2.4 The `msBayes` model

Let Y be the number of population pairs, k_i be the number of loci sampled for the i^{th} population pair, and K be the total number of unique sampled loci. Let $\mathbf{X} = \{X_{1,1}, \ldots, X_{Y,k_Y}\}$ represent the data, which is a vector of multiple sequence alignments, where $X_{i,j}$ is the alignment of the j^{th} locus sampled for the i^{th} population pair. The joint posterior distribution of

the model implemented in `msBayes` is given by

$$
\begin{aligned}
&f(\mathbf{G}, \Psi, \boldsymbol{\tau}, \boldsymbol{\theta_A}, \boldsymbol{\theta_{D1}}, \boldsymbol{\theta_{D2}}, \alpha, \boldsymbol{\upsilon}, \boldsymbol{\tau_B}, \boldsymbol{\zeta_{D1}}, \boldsymbol{\zeta_{D2}}, \mathbf{m} \,|\, \mathbf{X}, \boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\nu}) \\
&= \frac{1}{f(\mathbf{X})} f(\Psi) f(\boldsymbol{\tau} \,|\, \Psi) f(\alpha) \left[\prod_{i=1}^{Y} f(\theta_{A,i}) f(\theta_{D1,i}, \theta_{D2,i}) f(\tau_{B,i}) f(\zeta_{D1,i}) f(\zeta_{D2,i}) f(m_i) \right. \\
&\left. \prod_{j=1}^{k_i} f(X_{i,j} \,|\, G_{i,j}, \phi_{i,j}) f(G_{i,j} \,|\, \tau_i, \theta_{A,i}, \theta_{D1,i}, \theta_{D2,i}, \rho_{i,j}, \nu_{i,j}, \upsilon_j, \tau_{B,i}, \zeta_{D1,i}, \zeta_{D2,i}, m_i) \right] \left[\prod_{j=1}^{K} f(\upsilon_j \,|\, \alpha) \right],
\end{aligned}
\tag{1.1}
$$

where $\mathbf{G} = \{G_{1,1}, \ldots, G_{Y,k_Y}\}$ are the gene trees upon which each $X_{i,j}$ evolved according to the HKY85 substitution model parameters $\phi_{i,j}$. The HKY85 model parameters for each alignment reside in vector $\boldsymbol{\phi} = \{\phi_{1,1}, \ldots, \phi_{Y,k_Y}\}$, and are fixed constants provided by the user. $\boldsymbol{\tau}$ is the vector of times, in coalescent units, when the populations of each pair diverged, $\{\tau_1, \ldots, \tau_Y\}$. Ψ is the hyper-parameter controlling the number of unique τ within $\boldsymbol{\tau}$. $\boldsymbol{\theta_A} = \{\theta_{A,1}, \ldots, \theta_{A,Y}\}$ is the vector of θ parameters for the ancestral population of each population pair. $\boldsymbol{\theta_{D1}} = \{\theta_{D1,1}, \ldots, \theta_{D1,Y}\}$ is the vector of θ parameters for the 1^{st} descendant population of each pair; $\boldsymbol{\theta_{D2}} = \{\theta_{D2,1}, \ldots, \theta_{D2,Y}\}$ is the same for the 2^{nd} descendant population of each pair. $\boldsymbol{\rho} = \{\rho_{1,1}, \ldots, \rho_{Y,k_Y}\}$ and $\boldsymbol{\nu} = \{\nu_{1,1}, \ldots, \nu_{Y,k_Y}\}$ are vectors of θ-scaling constants provided by the user (Table 1.1). Furthermore, there are locus-specific θ-scaling parameters in the vector $\boldsymbol{\upsilon} = \{\upsilon_1, \ldots, \upsilon_K\}$. α is the shape parameter of the gamma prior distribution on each υ. $\boldsymbol{\zeta_{D1}} = \{\zeta_{D1,1}, \ldots, \zeta_{D1,Y}\}$ is the vector of θ-scaling parameters that determine the magnitude of the bottleneck in the 1^{st} descendant population of each population pair, whereas $\boldsymbol{\zeta_{D2}} = \{\zeta_{D2,1}, \ldots, \zeta_{D2,Y}\}$ is the same for the 2^{nd} descendant population of each pair; for both descendants of each population pair, the bottleneck begins immediately after divergence in forward-time. $\boldsymbol{\tau_B} = \{\tau_{B,1}, \ldots, \tau_{B,Y}\}$ is the vector of the proportions of time between present and τ when the bottleneck ends for both populations in each pair; after which the populations grow exponentially to present. $\mathbf{m} = \{m_1, \ldots, m_Y\}$ is the vector of symmetric migration rates between the descendant populations of each pair.

Prior terms of Equation (1.1)

The prior terms of Equation (1.1) within the product over population pairs include $f(\theta_{A,i})$, $f(\tau_{B,i})$, $f(\zeta_{D1,i})$, $f(\zeta_{D2,i})$, and $f(m_i)$. For each pair, these are independently and identically distributed (*iid*) as $\theta_A \sim U(a_\theta, b_{\theta_A})$; $\tau_B \sim U(0, 0.95)$; $\zeta_{D1} \sim U(0.01, 1)$; $\zeta_{D2} \sim U(0.01, 1)$; and $m \sim U(0, b_m)$. Let us further denote the mean of the two descendant populations of the i^{th} pair as $\theta_{D,i}$, which, for each pair, is *iid* as $\theta_D \sim U(a_\theta, b_{\theta_D})$. The prior on θ of the 1^{st} and 2^{nd} descendant population of the i^{th} pair ($f(\theta_{D1,i}, \theta_{D2,i})$) is then $\theta_{D1,i}, \theta_{D2,i} \sim Dirichlet(1,1) \times 2\theta_{D,i}$.

The terms in Equation (1.1) outside of the product over population pairs include the hyper-prior probability distributions $f(\Psi)$, $f(\alpha)$, and $f(\boldsymbol{\tau} \,|\, \Psi)$. The prior on Ψ is uniformly distributed on the integers 1 to Y. The prior on α is $\alpha \sim U(1, 20)$. If we let $\mathbf{T} = \{T_1, \ldots, T_\Psi\}$ be the vector of the Ψ unique divergence times, then $f(\boldsymbol{\tau} \,|\, \Psi) = f(\mathbf{T} \,|\, \Psi) f(\boldsymbol{\tau} \,|\, \mathbf{T})$. Each T within $\mathbf{T}$ is *iid* as $T \sim U(0, b_\tau)$. Each T is placed in $\boldsymbol{\tau}$ once, and the remaining $Y - \Psi$ slots within $\boldsymbol{\tau}$ are populated by randomly drawing from $\mathbf{T}$ with replacement. If we let $x_1, \ldots, x_\Psi$ denote the number of times each $T_1, \ldots, T_\Psi$ is selected for the $Y - \Psi$ slots, then the probability mass function is

$$f(\boldsymbol{\tau} \,|\, \mathbf{T}) = f(x_1, \ldots, x_\Psi; Y - \Psi, p_1, \ldots, p_\Psi) = \frac{Y - \Psi!}{x_1! \cdots x_\Psi!} p_1^{x_1} \cdots p_\Psi^{x_\Psi}, \tag{1.2}$$

where $p_1 = p_2 = \ldots = p_\Psi = 1/\Psi$ and $\sum_{i=1}^{\Psi} x_i = Y - \Psi$.

The τ parameters are in coalescent units relative to a constant reference population size, θ_C/μ, where $\theta_C = b_{\theta_D}/2$. We denote these coalescent units as $4N_C$ generations. Thus, the τ within $\boldsymbol{\tau}$ are proportional to real time, and can be converted to the number of generations of the reference population, τ_G, by assuming a mutation rate, μ, and using

$$\tau_G = \tau \times \frac{\frac{b_{\theta_D}}{2}}{\mu}. \tag{1.3}$$

Note, for each τ within $\boldsymbol{\tau}$ to be on the same scale of $4N_C$ generations, and thus comparable, `msBayes` assumes the relative mutation rates among the populations are fixed and known. The relative rates are fixed according to the values in $\boldsymbol{\nu}$ provided by the user. To get the divergence times into units proportional to the realized population size, and thus the expected number of mutations, `msBayes` scales the divergence times for each locus of each population pair before simulating data, creating the vector $\mathbf{t} = \{t_{1,1}, \ldots, t_{Y,k_Y}\}$. For the j^{th} locus of the i^{th} pair, τ_i is scaled by

$$t_{i,j} = \tau_i \times \frac{\theta_C}{\theta_{D,i}\rho_{i,j}}. \tag{1.4}$$

The user-defined θ-scaling constants in $\boldsymbol{\rho}$ can thus be used to account for known differences in ploidy among the loci and/or differences in generation times among the taxa.

`msBayes` allows for intra-locus recombination, which, for simplicity, is not included in Equation (1.1). If the intra-locus recombination rate, r, is allowed to be non-zero, another prior $f(r)$ would be outside the product over population pairs in Equation (1.1), and it would be distributed as $r \sim U(0, b_r)$.

The term within the product over the K unique loci is the prior probability density of the θ-scaling parameter of the j^{th} locus, $f(\upsilon_j \,|\, \alpha)$. Each of these parameters is *iid* as $\upsilon \sim \Gamma(\alpha, 1/\alpha)$.

We use Θ to denote all of the parameters of the `msBayes` model, and $f(\Theta)$ to represent the joint prior probability distribution of the model (Table 1.1).

Likelihood terms of Equation (1.1)

For the j^{th} locus of the i^{th} population pair, the term $f(X_{i,j} \,|\, G_{i,j}, \phi_{i,j})$ is the probability of the sequence alignment given a gene tree and HKY85 parameters, or the Felsenstein likelihood (Felsenstein, 1981). If the intra-locus recombination rate is allowed to be non-zero, Equation (1.1) would require another product over the columns of each sequence alignment

to allow sites to have different genealogies. The term $f(G_{i,j} \mid \ldots)$ is the probability of the gene tree under a multi-population coalescent model where the ancestral population of constant size $\theta_{A,i}\rho_{i,j}\nu_{i,j}\upsilon_j$ diverges at time τ_i into two descendant populations of constant size $\theta_{D1,i}\rho_{i,j}\nu_{i,j}\upsilon_j\zeta_{D1,i}$ and $\theta_{D2,i}\rho_{i,j}\nu_{i,j}\upsilon_j\zeta_{D2,i}$ that exchange migrants at symmetric rate m_i. After time $\tau_{B,i} \times \tau_i$ they grow exponentially to present size $\theta_{D1,i}\rho_{i,j}\nu_{i,j}\upsilon_j$ and $\theta_{D2,i}\rho_{i,j}\nu_{i,j}\upsilon_j$, respectively. Lastly, $f(\mathbf{X})$ is the marginal likelihood of the model, or the probability of the data.

1.2.5 The ABC implementation of the `msBayes` model

`msBayes` does not estimate the posterior in Equation (1.1). Rather, it distills the alignments $\mathbf{X} = \{X_{1,1}, \ldots, X_{Y,k_Y}\}$ into vectors of summary statistics $\mathbf{S}^* = \{S^*_{1,1}, \ldots, S^*_{Y,k_Y}\}$ and estimates the approximate joint posterior distribution

$$
\begin{aligned}
&f(\mathbf{G}, \Psi, \boldsymbol{\tau}, \boldsymbol{\theta_A}, \boldsymbol{\theta_{D1}}, \boldsymbol{\theta_{D2}}, \alpha, \boldsymbol{\upsilon}, \boldsymbol{\tau_B}, \boldsymbol{\zeta_{D1}}, \boldsymbol{\zeta_{D2}}, \mathbf{m} \mid B_\epsilon(\mathbf{S}^*), \boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\nu}) \\
&= \frac{1}{f(B_\epsilon(\mathbf{S}^*))} f(\Psi) f(\boldsymbol{\tau} \mid \Psi) f(\alpha) \left[\prod_{i=1}^{Y} f(\theta_{A,i}) f(\theta_{D1,i}, \theta_{D2,i}) f(\tau_{B,i}) f(\zeta_{D1,i}) f(\zeta_{D2,i}) f(m_i) \right. \\
&\left. \prod_{j=1}^{k_i} f(B_\epsilon(S^*_{i,j}) \mid G_{i,j}, \phi_{i,j}) f(G_{i,j} \mid \tau_i, \theta_{A,i}, \theta_{D1,i}, \theta_{D2,i}, \rho_{i,j}, \nu_{i,j}, \upsilon_j, \tau_{B,i}, \zeta_{D1,i}, \zeta_{D2,i}, m_i) \right] \left[\prod_{j=1}^{K} f(\upsilon_j \mid \alpha) \right],
\end{aligned}
\tag{1.5}
$$

where $B_\epsilon(\mathbf{S}^*)$ is the multi-dimensional Euclidean space around the vector of observed summary statistics, the radius of which is the tolerance ϵ.

To estimate this approximate posterior, `msBayes` uses an ABC rejection algorithm followed by regression adjustment. The first step of the algorithm is to draw n samples from the joint prior, $f(\Theta)$. In the case of one locus per population pair, `msBayes` draws a sample from $f(\Theta)$ by (1) drawing Ψ from the integers 1 to Y; (2) drawing Ψ divergence times from $U(0, b_\tau)$ in units of $4N_C$ generations, to get $\mathbf{T} = \{T_1, \ldots, T_\Psi\}$; (3) randomly assigning each T to $\boldsymbol{\tau}$ once, and filling the remaining $Y - \Psi$ slots within $\boldsymbol{\tau}$ by randomly drawing from $\mathbf{T}$ with replacement, to get $\boldsymbol{\tau} = \{\tau_1, \ldots, \tau_Y\}$; (4) drawing values of the demographic parame-

ters θ_A, θ_{D1}, θ_{D2}, τ_B, ζ_{D1}, ζ_{D2}, and m for each population pair from their respective prior distributions; (5) scaling each τ to get $\mathbf{t} = \{t_{1,1}, \ldots, t_{Y,k_Y}\}$ via Equation (1.4); (6) simulating sequence data for each population pair according to the multi-population coalescent model described above; and (7) calculating population genetic summary statistics S for each pair from the simulated sequence matrix. The result of one draw from the prior is the parameter vector Λ and vector of summary statistics $\mathbf{S} = \{S_1, \ldots, S_Y\}$. For Λ, `msBayes` reports Ψ and three summary statistics calculated from $\boldsymbol{\tau}$: the mean ($E(\tau)$), variance ($Var(\tau)$), and dispersion index ($\Omega = Var(\tau)/E(\tau)$) (Hickerson et al., 2006; Huang et al., 2011). After repeating this process n times, we have a sample of parameter vectors, $\mathcal{P}_{f(\Theta)} = \{\Lambda_1, \ldots, \Lambda_n\}$, randomly drawn from $f(\Theta)$, and the associated vectors of summary statistics, $\mathbb{S} = \{\mathbf{S_1}, \ldots, \mathbf{S_n}\}$.

The vector $\mathbf{S}^*$ contains the same summary statistics calculated from the observed sequence data. During the rejection step, only the samples from $\mathcal{P}_{f(\Theta)}$ with $\mathbf{S}$ that fall within the Euclidean space $B_\epsilon(\mathbf{S}^*)$ are retained. We denote the set of retained samples as $\mathcal{P}_\epsilon$. Regression techniques are then used to adjust $\mathcal{P}_\epsilon$ for variation in the probability of the data across the retained parameter sample space (Beaumont et al., 2002; Leuenberger and Wegmann, 2010; Blum and François, 2009). The result is an estimate of the approximate posterior, which we denote $\mathcal{P}_{f(\Theta|B_\epsilon(\mathbf{S}^*))}$.

1.2.6 Estimating the pattern of divergence times

To estimate the temporal pattern of divergences among the 22 population pairs, we used a modified version of `msBayes` v20100519 (Huang et al., 2011). In the midst of our work, we identified a bug in version 20100519 that mis-specified the prior on the θ_A parameters, which we subsequently corrected. This error has also been corrected in version 20120222 of `msBayes` (full details in the Appendix).

Specifying and simulating the joint prior

To use `msBayes`, one must specify b_τ, a_θ, b_{θ_D}, b_{θ_A}, b_m, and b_r to control the prior distributions $\tau \sim U(0, b_\tau)$, $\theta_D \sim U(a_\theta, b_{\theta_D})$, $\theta_A \sim U(a_\theta, b_{\theta_A})$; $m \sim U(0, b_m)$; and $r \sim U(0, b_r)$. Given our island system and mitochondrial data, we assumed no migration ($b_m = 0$) and no recombination ($b_r = 0$). Currently, `msBayes` only offers the continuous uniform distribution to represent *a priori* knowledge about parameters of θ and τ. Thus, broad prior distributions must be used to avoid assigning zero probability density to plausible regions of parameter space. In our case, we must choose prior distributions that span the range of possible values for the 66 θ and 22 τ parameters. Specifically, we chose prior settings of $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$. We chose the θ settings to assure we spanned the 30 empirical estimates of θ calculated from independent data. We chose the prior on τ to represent the large *a priori* uncertainty about the divergence times for all 22 pairs of populations. Two of our population pairs (*Crocidura negrina-C. panayensis* and *Cyrtodactylus gubaot-C. sumuroi*) represent distinct species, and the species-level taxonomy of many vertebrate groups in the Philippines often masks deeply divergent cryptic lineages (Brown et al., 2008; Siler and Brown, 2010; Siler et al., 2011a,b, 2012; Welton et al., 2010; Linkem et al., 2010; Brown and Stuart, 2012; Esselstyn et al., 2012). Given the precedents for cryptic diversity in the Philippines, the long geological history of the archipelago (Brown and Diesmos, 2009; Yumul et al., 2008), and the large amount of uncertainty regarding the mutation rates and generation times of the taxa examined, we chose $b_\tau = 20$ to avoid giving plausible divergence times zero probability. Assuming $\mu = 2 \times 10^{-8}$ and applying Equation (1.3), this prior translates to $\tau_G \sim U(0, 5 \times 10^7)$ in generations.

We used the `msbayes.pl` script to generate $\mathcal{P}_{f(\Theta)}$ with 10^7 random samples from $f(\Theta)$, using the ML estimates of the HKY85 model parameters for each of the population pairs and assuming μ is equal across all taxa. Each $\mathbf{S}$ contained π (Tajima, 1983), θ_W (Watterson, 1975), π_{net} (Takahata and Nei, 1985), and $SD(\pi - \theta_W)$ (Tajima, 1989) for each of the 22 population pairs.

Rejection sampling and regression adjustment

We used the `acceptRej.pl` script of the `msBayes` package and ABCtoolbox (v1.1; Wegmann et al., 2010) to estimate $\mathcal{P}_\epsilon$ by specifying ϵ so that 1000 samples from $\mathcal{P}_{f(\Theta)}$ were within $B_\epsilon(\mathbf{S}^*)$ and thus retained. ABCtoolbox and `msBayes` produced identical $\mathcal{P}_\epsilon$. We used two regression methods to adjust $\mathcal{P}_\epsilon$ and estimate the approximate posterior, $\mathcal{P}_{f(\Theta|B_\epsilon(\mathbf{S}^*))}$: (1) the weighted, local-linear regression (ABC_{LLR}) adjustment of Beaumont et al. (2002) as implemented in `msBayes` and (2) the general linear model (ABC_{GLM}) regression adjustment of Leuenberger and Wegmann (2010) as implemented in ABCtoolbox. To avoid additional notation, we include the multinomial logistic regression adjustment of Ψ used by `msBayes` under ABC_{LLR}. For both methods we used the same set of summary statistics as for the rejection step.

Vetting prior sample size and PAIC-specific analyses

To assess whether 10^7 samples from $f(\Theta)$ were sufficient, we split the prior samples into two sets of 5×10^6 and repeated the rejection sampling and ABC_{GLM}-regression adjustment using each subset. To compare the results between the two different PAICs, we repeated the ABC methods above on the nine population pairs from the Greater Negros-Panay PAIC and the 13 population pairs from the Greater Mindanao PAIC (Table S5.3) separately; we generated 5×10^6 prior samples for these PAIC-specific analyses.

Reducing summary statistics to PLS components

In practice, using too many summary statistics may introduce noise into ABC estimates (Joyce and Marjoram, 2008; Wegmann et al., 2009). In our case, each $\mathbf{S}$ contains 88 summary statistics (the four default `msBayes` summary statistics for each of the 22 population pairs). To reduce the dimensionality of $\mathbb{S}$, we transformed the 88 summary statistics into 10 partial least squares (PLS) orthogonal components, following an initial Box-Cox transformation of the statistics (Wegmann et al., 2009). We used the `find_pls.r` script of ABCtoolbox to

define the PLS components using all 10^7 samples within $\mathbb{S}$. Using these PLS definitions, we reduced $\mathbb{S}$ and $\mathbf{S}^*$ to $\mathbb{S}_{PLS}$ and $\mathbf{S}^*_{PLS}$ with 10 PLS components. Hereafter, we will refer to the prior sample containing the original 88 summary statistics as $\mathbb{S}_{stats}$ to distinguish it from $\mathbb{S}_{PLS}$. We repeated all the rejection-sampling and regression adjustment procedures discussed above using $\mathbb{S}_{PLS}$ and $\mathbf{S}^*_{PLS}$. Accommodating the PLS components during the rejection sampling and ABC_{LLR} adjustment in `msBayes` required modifications to the `acceptRej.pl` and `acceptRej.r` scripts provided with the package; the modified scripts are available at `https://github.com/joaks1/msbayes-hack`.

1.2.7 Assessing the performance and power of `msBayes`

What does "simultaneous" mean?

The PAIC model predicts a pattern of recent and clustered divergences among co-distributed taxa (i.e., groups of divergence times associated with Pleistocene glacial cycles). Using `msBayes` to test this prediction assumes the method can reliably discriminate recent clustered divergences from random divergences. There have not been any simulation-based assessments of the power of `msBayes` to detect variation in divergence times. We simulated 1000 pseudo-replicate datasets (i.e., 1000 Λ with associated $\mathbf{S}$) with τ for each of the 22 population pairs (i.e., $\Psi = 22$) randomly drawn from a uniform distribution, $U(0, \tau_{max})$, where τ_{max} was set to: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 2.0, and 3.0, in $4N_C$ generations. We used the `msbayes.pl` script to generate the pseudo-replicates with all other settings the same as when we generated $\mathcal{P}_{f(\Theta)}$.

For each of the 17,000 simulated datasets, we repeated the same ABC inference procedures that we used for the empirical data. Specifically, we used $\mathbb{S}_{stats}$, ABC_{GLM}, and 5×10^6 samples from $f(\Theta)$. Using version 20100519 of `msBayes` (i.e., pre-bug fix), we explored all of the ABC methods used for the empirical data (ABC_{LLR} versus ABC_{GLM} and $\mathbb{S}_{stats}$ versus $\mathbb{S}_{PLS}$) as well as 2×10^6 versus 10^7 samples from the prior. All of these methods performed similarly (see Appendix). The error in version 20100519 of `msBayes` enforces a lower bound

of 0.01 on the prior for θ_A (see Appendix for specifics). This hard-coded bound was the same for simulations generating both the pseudo-observed datasets and the prior sample (i.e., the prior was correct). Thus, the bug should have little effect on the relative performance of the methods.

1.2.8 Assessing prior sensitivity of `msBayes`

Our simulation-based analyses reveal a bias toward inferring clustered divergences (see Results), which could be caused by broad priors on θ and τ (Lindley, 1957). We studied this in two ways. First, we quantified the behavior of `msBayes` under the ideal conditions where the prior distributions are correct. To do this, we simulated 100,000 datasets by drawing parameter values from the prior ($\tau \sim U(0, 10)$, $\theta_D \sim U(0.0001, 0.05)$, and $\theta_A \sim U(0.0001, 0.025)$) and analyzed these datasets using 2×10^6 samples drawn from the same prior distributions. For all these simulations, we used the default distribution on Ψ (i.e., discrete uniform from 1 to Y). To greatly reduce computational time, these simulations used a single 1000 bp locus sampled from 10 individuals from 10 population pairs. We were particularly interested in the posterior probability of a single divergence time for each of these simulated datasets. So, we assigned the 100,000 estimates of the posterior probability of one divergence event (i.e., $p(\Psi = 1|B_\epsilon(\mathbf{S}^*))$) to 20 bins of width 0.05, and plotted the estimated $p(\Psi = 1|B_\epsilon(\mathbf{S}^*))$ of each bin against the proportion of replicates in that bin with a true value of $\Psi = 1$ (Huelsenbeck and Rannala, 2004a). We repeated this exercise using the alternative criterion for one divergence event, $\hat{\Omega} < 0.01$ (Hickerson et al., 2006).

Our second evaluation of prior sensitivity attempted to assess the performance of `msBayes` under optimal real-world conditions. In this case, the priors used are not known to be correct, but represent the narrowest possible prior distributions (i.e., priors informed by the data). We repeated the simulation-based power analyses outlined above using two additional, highly informative prior settings: (1) $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$; and (2) $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$. For both prior settings,

we generated 5×10^6 samples. Using the informed θ priors, we simulated an additional 17,000 pseudo-observed datasets with divergence times randomly drawn from the same series of uniform distributions (i.e., τ_{max}) described above. The priors on θ match the range of θ estimates from our empirical data (Tables S5.4 and S5.5). Assuming $\mu = 2 \times 10^{-8}$, the priors on τ translate to $\tau_G \sim U(0, 1 \times 10^7)$ and $\tau_G \sim U(0, 5 \times 10^6)$ in units of generations of the reference population, respectively. If we assume one generation per year, the latter prior is very similar to our range of gene divergence estimates (0.2–4 mybp; Fig. S5.2), and thus, given the uncertainty in generation times and mutation rates for the taxa we examined, is likely too narrow (i.e., excluding plausbile values).

We must emphasize that we only use such informed priors here to represent the narrowest possible priors for a real-world application of `msBayes`. Basing priors on estimates from the data is not a fully Bayesian statistical procedure, and we are not advocating such an approach. Whereas empirical Bayes, also known as maximum marginal likelihood, is a commonly used statistical framework for point estimation, empirical Bayesian estimates of the posterior distribution of parameters are too narrow and often inappropriately shaped and off-center (Morris, 1983; Laird and Louis, 1987; Carlin and Gelfand, 1990). This is because they fail to account for the uncertainty in estimating the prior. Many post-hoc correction methods have been proposed for estimating frequentist-like confidence intervals from empirical Bayesian estimates of the posterior distribution of parameters (Morris, 1983; Laird and Louis, 1987, 1989; Carlin and Gelfand, 1990; Hwang et al., 2009), but none are implemented in `msBayes`, and none would correct model choice estimates (e.g., posterior probabilities of models and Bayes factors).

1.2.9 A note on coalescent units

Because our empirical data is comprised of mitochondrial sequences, our estimates of $E(\tau)$ and Ω are in coalescent units of N_C generations. However, the coalescent units of our simulated pseudo-observed data are arbitrary and, via Equation (1.3), can be converted into

generations by assuming μ (i.e., the conversion is independent of the inheritance constant of the coalescent units). When reporting our simulation results in coalescent units, we write "$4N_C$" generations, because most users of `msBayes` will be using diploid, biparentally inherited loci.

1.3 Results

1.3.1 Estimates of gene divergences and θ

Our estimates of the gene divergence times range from approximately 0.2 to 4 mybp, with 16 of the 22 posterior mean estimates within the past million years (Fig. S5.2). Our estimates of θ from independent population samples range from 0.0011 to 0.0181 (Table S5.4); our estimates from the 44 populations we analyzed in `msBayes` had a greater range of 0.0003 to 0.0381 (Table S5.5).

1.3.2 Empirical `msBayes` estimates under broad priors

When using prior settings of $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$, our `msBayes` results strongly support one recent divergence event for all 22 population pairs, regardless of the method of post-sampling regression (ABC_{LLR} or ABC_{GLM}) or summary statistic matrix ($\mathbb{S}_{stats}$ or $\mathbb{S}_{PLS}$) used (Fig. S5.3 and Table 1.2). Consistent with one divergence event, all methods yield estimates of the dispersion index of population divergence times (Ω) of essentially zero (Table 1.2). $\hat{\Omega} < 0.01$ is commonly used as a criterion for one divergence time shared by all pairs of populations (Hickerson et al., 2006).

Estimates of the time of the divergence event ($\hat{E(\tau)}$) range from approximately 0.04–0.1 coalescent units ago (Table 1.2). Assuming $\mu = 2 \times 10^{-8}$ and one generation per year, this translates to 100,000–250,000 years ago, consistent with Pleistocene-driven diversification. We used simulations to reject the possibility that these results can be explained by a model of no divergence (i.e., panmixia between the populations of each pair; see Appendix).

ABC_{LLR} and ABC_{GLM} estimates of $E(\tau)$ are almost identical, regardless of whether using $\mathbb{S}_{stats}$ or $\mathbb{S}_{PLS}$. Both regression methods had larger confidence intervals when using $\mathbb{S}_{PLS}$ (Table 1.2). Estimates based on 5×10^6 and 10^7 prior samples are very similar, regardless of whether $\mathbb{S}_{stats}$ or $\mathbb{S}_{PLS}$ are used (Table 1.2). Estimates from the Greater Mindanao and Negros-Panay PAICs were similar to the combined analyses (Table 1.2).

1.3.3 Simulation-based assessment of "simultaneous"

Accuracy and precision of estimates with broad priors

The precision of the ABC estimates of $E(\tau)$ and Ω is low, especially when the true divergence times are more recent (Figs. 1.1 and S5.4). Also, `msBayes` is less accurate and precise in estimating Ω (Fig. 1.1) than $E(\tau)$ (Fig. S5.4). When τ_{max} is less than 0.9 coalescent units, `msBayes` tends to underestimate Ω, whereas it tends to overestimate Ω when τ_{max} is 1.0 or greater (Fig. 1.1). From our simulation results using version 20100519 of `msBayes` (pre-bug fix), all combinations of summary statistics and regression-adjustment methods are inaccurate for estimating Ω across the τ_{max} we simulated (Figs. S5.18–5.21).

Power of `msBayes`

The power of `msBayes` to detect variation in divergence times is low at phylogeographic time scales (Figs. 1.2 and 1.3). If we judge the procedure by conditions that lead to $\leq 5\%$ of the simulation replicates estimating one divergence event, then we find that `msBayes` is unable to reject one divergence event based on $\hat{\Psi}$ when τ_{max} is less than 1.3 coalescent units (i.e., $5.2N_C$ generations; Fig. 1.2). This translates to `msBayes` inferring a single divergence event more than 5% of the time when the 22 population pairs diverged randomly over the past 3 million generations (assuming $\mu = 2 \times 10^{-8}$).

`msBayes` does better based on $\hat{\Omega}$, but still cannot reject one divergence when τ_{max} is less than 0.8 coalescent units (i.e., $3.2N_C$ generations, or 2 million generations assuming $\mu = 2 \times 10^{-8}$; Fig. 1.3). This lack of power was not due to an overly stringent threshold

(0.01). The true values of Ω from the simulations under all τ_{max} are consistent with multiple divergence times (i.e., greater than 0.01; Fig. S5.5). Thus, based on the criterion $\hat{\Omega} < 0.01$, `msBayes` would have rejected one divergence event for all the τ_{max} values we simulated if it was able to accurately estimate Ω. The lack of power is due to the bias of `msBayes` to underestimate Ω at recent divergence times (Fig. 1.1).

We can also assess power of the method by examining what conditions lead to $\leq 5\%$ of the simulation replicates estimating strong support for $\Psi = 1$. We use a Bayes factor of greater than 10 for the one divergence model compared to all other models ($BF_{\Psi=1,\Psi\neq1} > 10$) as a threshold for strong support (Jeffreys, 1935, 1961). According to this test, the method cannot reject one divergence when τ_{max} is less than 1.3 coalescent units (Fig. 5.6). Thus, assuming $\mu = 2 \times 10^{-8}$, `msBayes` strongly supports one divergence event more than 5% of the time when divergence times are random over the past 3 million generations. If we increase our threshold for strong support to a posterior probability greater than 0.95 ($BF_{\Psi=1,\Psi\neq1} > 399$), `msBayes` is still unable to reject one divergence event when τ_{max} is less than 0.8 coalescent units (2 million generations).

Using the criteria above based on the inference of one divergence event (or strong support for one divergence event) is quite generous, because an inference of $\hat{\Omega} > 0.01$ or $\hat{\Psi} > 1$ is not equivalent to success of the method. This is clearly illustrated by Figure 1.2. `msBayes` infers highly clustered divergence times across all of the τ_{max} we simulated. For example, even when divergence times are random over the past 3 coalescent units (7.5 million generations), the most probable inference of `msBayes` is still only two divergence events (i.e., $\hat{\Psi} = 2$).

Based on our simulation results using version 20100519 of `msBayes`, the power to detect variation in divergence times was low under all combinations of summary statistics, regression-adjustment methods, and prior sample sizes we explored (Figs. S5.22–5.28 and S5.33–5.35).

1.3.4 Prior sensitivity of `msBayes`

Simulation results

Under the ideal conditions when the priors are correct, `msBayes` provides reasonable estimates of the posterior probability of one divergence event (Fig. 1.4). Based on Ψ and Ω results from $\mathcal{P}_\epsilon$ (i.e., unadjusted estimates), `msBayes` overestimates the posterior probability of one divergence event when the true probability is less than ≈ 0.6, and slightly underestimates it when the true probability is greater than ≈ 0.6 (Fig. 1.4A&B). However, regression adjustment of Ω using both ABC_{LLR} and ABC_{GLM} causes extreme underestimation of the posterior probability of one divergence event (Fig. 1.4D&F). Regression adjustment of Ψ has the same affect, but less extreme (Fig. 1.4C&E). Thus, Ψ is a better estimator of the posterior probability of one divergence event than Ω (Fig. 1.4C–F). Note that the adjustment of Ψ in `msBayes` (using multinomial logistic regression) failed for approximately 2% (2043) of the simulations, almost all of which had a true Ψ of one. Based on the unadjusted estimates, these failures would have been underestimates of the posterior probability of $\Psi = 1$. Thus, the plotted ABC_{LLR} results (Fig. 1.4C) are skewed due to these failed replicates, and should look more like the ABC_{GLM} results (Fig. 1.4E). Overall, when the priors are correct, regression-adjusted estimates of the posterior probability of one divergence event are downward biased. This is consistent with our observation of the switch in bias of Ω estimates; as τ_{max} becomes more similar to the prior, the bias switches from upward to downward (Fig. 1.1).

When the priors are informed by the data (i.e., narrower than possible under real-world applications of `msBayes`), the method's power is similar to using broad priors. Based on estimates of Ψ, `msBayes` cannot reject one divergence event when τ_{max} is less than 1.3 coalescent units (1.3 million generations, assuming $\mu = 2 \times 10^{-8}$) when prior settings are $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$, and 1.2 coalescent units when the prior settings are $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$ (Figs. S5.10 and

S5.11). Also, under both informed prior settings, `msBayes` still infers highly clustered divergences for all the τ_{max} we simulated (Figs. S5.10 and S5.11). The bias of `msBayes` to underestimate Ω at recent divergence times remains when using the informed priors (Figs. S5.8 and S5.9). As a result, the lack of power to detect variation in divergence times based on Ω estimates also remains (Figs. S5.12 and S5.13). Specifically, the method cannot reject a single divergence event when τ_{max} is less than 0.9 coalescent units (900,000 generations, assuming $\mu = 2 \times 10^{-8}$) when the prior settings are $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$, and 0.8 coalescent units when the prior settings are $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$ (Figs. S5.12 and S5.13).

Empirical results

The results of `msBayes` analyses of the empirical data are very sensitive to the different prior settings we explored (Table 1.3). The estimates of Ω and $E(\tau)$ vary by over two and one order of magnitude, respectively, and the estimated 95% highest posterior density intervals do not overlap between some of the analyses. Our estimate of the posterior probability of one divergence event is essentially one when using appropriately broad priors, but is nearly zero with narrow priors (Table 1.3). Thus, we find strong support for different biogeographic scenarios (i.e., one divergence versus multiple divergences) under different prior distributions. However, under the informed priors, `msBayes` still infers highly clustered divergence times ($\hat{\Psi} = 2$; Table 1.3), suggesting the estimates under all three prior settings may suffer from the biases revealed by our simulation results.

1.4 Discussion

1.4.1 Power and bias issues

Given the strong support for either one or two shared recent divergence(s) among the 22 taxon pairs (Tables 1.2 and 1.3), it would have been easy to accept the empirical results

of `msBayes` as evidence for climate-driven diversification across the Philippine Archipelago. However, our simulations demonstrate a bias in the `msBayes` model toward inferring clustered divergences among population pairs when the divergences are random and relatively recent. We found `msBayes` always inferred temporally clustered divergences even when the taxon pairs diverged randomly over the past $12N_C$ generations (Fig. 1.2L). To put this in real time, assuming a mutation rate of 2×10^{-8} and a one year generation time, `msBayes` consistently infers an interesting biogeographical scenario when the taxon pairs diverged randomly over the past 7.5 million years. Also, `msBayes` will often ($> 5\%$ of the time) infer the extreme case of *one* divergence event with strong support when the taxon pairs diverged randomly over the past $3.2N_C$ generations (2 million years) (Fig. 1.3 and S5.6).

The bias did not improve when using empirically guided prior distributions (Fig. S5.8, S5.9, S5.12,and S5.13). Thus, our results show that even when using priors that are more informative than could be expected in a real-world application of `msBayes`, the method tends to infer simultaneous divergence too frequently.

`msBayes` is primarily used for comparing shallow divergences (Hickerson et al., 2006; Huang et al., 2011), which, according to our results, is when the method can be the most misleading. Due to the stochasticity of coalescent and mutational processes, a large variance in genetic divergence is expected among recently co-diverged taxa. Thus, any inference method for estimating divergence times is expected to struggle when applied to recently diverged taxa (and over much of the range of parameters we simulated here). Ideally, a method should express uncertainty in the face of such large expected variance. However, our simulations indicate that `msBayes` often returns strong support for the spurious conclusion of one divergence event (Fig. S5.6).

These findings are worrisome, because co-divergence is often a biologically interesting result that is interpreted as evidence for a shared historical event or barrier. Our results suggest that any application of `msBayes` on recently diverged taxa is very likely to result in clustered divergences (Figs. 1.2, S5.10, and S5.11), and thus an "interesting" biogeographical

interpretation.

1.4.2 Prior sensitivity of empirical estimates

We find strong support for contradictory hypotheses (i.e., one versus multiple divergence events) and very different parameter estimates under the prior settings we explored (Table 1.3). This sensitivity is problematic, and introduces another challenge to interpreting the results of `msBayes` analyses. When different prior settings yield contrasting results, it is difficult to draw biologically meaningful conclusions.

The same sensitivity to the prior settings was not observed in our simulation-based results. This is not unexpected. All of the pseudo-observed data were generated by the `msBayes` model, with all of the parameters (except $\boldsymbol{\tau}$) drawn from the prior distributions (i.e., the priors were correct). As a result, changes in the priors for the parameters θ_D and θ_A were reflected in the simulated data. This, of course, is not true for our empirical data, where the parameter values are unknown. Thus, the empirical estimates are expected to be more variable under the different prior settings than the simulation results. Also, the empirical data likely violate assumptions of the `msBayes` model that are met by the simulated data, which could contribute to the prior sensitivity we observed with the empirical data.

The empirical estimates of Ψ switched from one to two between the broad and informed priors (Table 1.3). Such a switch occurred for some of the pseudo-replicate datasets simulated at larger τ_{max} values and subsequently analyzed under a prior of $\tau \sim U(0, 10)$ versus $\tau \sim U(0, 5)$ (Figs. S5.10 and S5.11). Thus, some of the prior sensitivity we observed in analyses of the empirical data is also evidenced in the simulation results.

Our correction of the error in version v20100519 of `msBayes` concerning the θ_A prior also had an impact on the results of the empirical analyses. Our Ω estimates vary as much as two-fold before and after this error was corrected (Table S5.7). Given the sensitivity of our results to this error and prior settings, and the biases revealed by our simulations, we suggest that previously published results of `msBayes` be treated with some caution.

1.4.3 Possible causes of bias

`msBayes` implements a model selection procedure by approximating the posterior probabilities for different divergence models. The total number of unique divergence time models, or the number of possible ways to assign Y taxon pairs into Ψ divergence time categories, is calculated by the Bell number (Bell, 1934)

$$B_Y = \sum_{\Psi=1}^{Y} \left[\frac{1}{\Psi!} \sum_{j=0}^{\Psi-1} (-1)^j \binom{\Psi}{j} (\Psi - j)^Y \right]. \tag{1.6}$$

The number of unique models is enormous, even with a moderate number of taxon pairs (4,506,715,738,447,323 unique divergence models when $Y = 22$). The model implemented in `msBayes` takes advantage of the independence of the τ parameters and empirical sample sizes, allowing exchangeability of the summary statistics simulated under the various empirical sample sizes (Hickerson et al., 2006; Huang et al., 2011). This allows the identity/order of each τ within $\boldsymbol{\tau}$ to be ignored, which greatly reduces the divergence model space (1002 divergence models when $Y = 22$). The total number of unique divergence time models reduces to the number of partitions of Y (i.e., the integer partition), or the number of ways you can write Y as a sum of positive integers when the order of the addends does not matter (Sloan, 2011a,b). The closed-form expression for the integer partition is beyond the scope of this paper (but see Malenfant, 2011), but functions written in python to calculate it are provided in the appendix.

The prior on divergence models

In `msBayes`, the prior on Ψ is discrete uniform from 1 to Y, but it is not uniform over all the possible divergence models. For example, in our case $\Psi = 1$ and $\Psi = 22$ both represent a single divergence model, whereas $\Psi = 6$ comprises 136 models. Generally, we can calculate

the prior probability of the of the ith unique divergence model of class Ψ by

$$f(M_{\Psi,i}) = f(\Psi)\frac{1}{A(Y,\Psi)}, \tag{1.7}$$

where $f(\Psi)$ is the prior probability of Ψ, and $A(Y,\Psi)$ is the number of partitions of Y into Ψ divergence times (i.e., the number of unique divergence models for a given Ψ; Appendix 1; Sloan, 2011b). In the case of `msBayes`, the prior on Ψ is uniform discrete, so Equation (1.7) simplifies to

$$f(M_{\Psi,i}) = \frac{1}{Y \times A(Y,\Psi)}. \tag{1.8}$$

The distribution of the number of divergence models across Ψ, and corresponding prior probability distribution of divergence time models is shown in Fig. 1.5. In our case, the $M_{\Psi=1,i}$ model is 136 times more probable than a $M_{\Psi=6,i}$ model, *a priori*. The combination of the prior distribution over $M_{\Psi,i}$ (Fig. 1.5B) and potentially small marginal likelihoods of models with large Ψ (discussed below) could create a strong bias towards models with small Ψ. An alternative prior that places a uniform probability across each possible divergence model, rather than across Ψ, would result in a lower posterior for the single divergence event scenario whenever $Y > 3$.

Decreasing marginal likelihoods with increasing Ψ

The preference for clustered divergence models (i.e., small Ψ) implies that the models with larger Ψ have lower marginal likelihoods. The marginal likelihood of a model is an integral over the entire parameter space of the likelihood weighted by the prior probability density. By using broad, uniform priors for each divergence time parameter, we force models with many distinct divergence times to integrate over a *much* larger parameter space. If most of the parameter space has low likelihood, the marginal likelihood will be small. For example, consider the comparison of the models with $\Psi = 1$ and $\Psi = 22$. In our analyses, each of

the 22 divergence time parameters of the $\Psi = 22$ model has a uniform prior from zero to 20 coalescent units. If most of the prior space of these divergence time parameters has low likelihood density, integrating over the vast 22-dimensional τ parameter space will result in a low marginal likelihood for the $\Psi = 22$ model compared to the model with only one τ parameter.

Bayesian parameter estimation is often relatively robust to the choice of prior, particularly if the prior is vague. But model selection in the Bayesian context can be strongly influenced by excessively broad priors on nuisance parameters (Lindley, 1957; Jeffreys, 1961). This sensitivity to priors should hold for fully Bayesian model selection and for ABC methods. It is not a weakness of the analysis paradigm, but merely an indication that the prior probability statements used for nuisance parameters must be carefully chosen.

If prior sensitivity in our analyses caused the bias toward small Ψ, then a simulation study in which all parameters are drawn from the prior distributions used in the analysis (Huelsenbeck and Rannala, 2004a) should result in good performance with respect to the posterior probability of different values of Ψ. Our results demonstrate that when the priors are correct (i.e., there is no model misspecification) `msBayes` tends to be biased in the opposite direction and underestimates the posterior probability of one divergence event (Fig.1.4C–F). This suggests the bias toward clustered divergences when the prior on τ is broader than the true underlying distribution is caused, at least in part, by broad uniform priors reducing the marginal likelihoods of models with more τ parameters (i.e., larger Ψ).

However, our results suggest that, in practice, the uniform prior distributions may never be narrow enough to obviate the bias of `msBayes` toward clustered divergences. Even when we tested uniform priors more informative than could be expected in practice, the bias remained (Figs. 1.1–1.3 and S5.8–5.13). Perhaps more flexible probability distributions (e.g., gamma or log-normal) could better represent prior knowledge about θ and τ, while not placing too much prior probability density in regions of parameter space with low likelihood. Exploring alternative prior distributions may improve performance.

Insufficient sampling of parameter values

It is also possible that the estimates of the posterior probability for models with large Ψ will be inaccurate because of insufficient sampling of parameter values. If the prior over Ψ is uniform, models with larger prior parameter space (i.e., larger Ψ) will be less densely sampled. It is unclear if this phenomenon would bias the analysis toward models with fewer τ parameters or merely lead to higher variance in the estimates of the posterior probabilities. If this phenomenon is causing the bias, we expect analyses to be sensitive to the number of samples drawn from the prior. From our empirical analyses, we do not see such sensitivity when comparing prior sample sizes of 5×10^6 and 10^7. Also, our simulation results were unaffected by prior sample sizes of 2×10^6, 5×10^6, and 10^7 (see Appendix for full details). Thus, it seems unlikely that insufficient sampling of parameter space contributed to the bias.

Problems with ABC model choice

Recent work by Robert et al. (2011) has shown that ABC will often be biased in model choice. When summary statistics are insufficient for discriminating among competing models, which is the case for most empirical applications, ABC can be an inconsistent estimator of the models' posterior probabilities and can thus prefer the wrong model. This can occur even when the summary statistics are sufficient for each model under consideration Robert et al. (2011). The magnitude of the bias caused by the insufficiency of the statistics across models is unknown in most empirical settings, so there is no possibility of a correction factor (Robert et al., 2011). Thus, this problem does not have an obvious fix.

Whether the bias we observed is due to theoretical shortcomings of ABC model choice, broad uniform priors on nuisance parameters, the prior on divergence models, or some combination of these factors requires further investigation.

1.4.4 General Recommendations

Despite the limitations we identified, we conclude that `msBayes` can be a useful comparative phylogeographic tool. For example, `msBayes` can be used to test biogeographic hypotheses that require low temporal precision to be differentiated (i.e., on the scale of millions of generations). Alternatively, given the bias toward clustered divergences, an inference of no co-divergence (e.g., Topp and Winker, 2008) is likely robust. Also, there are several avenues to explore that might mitigate the problems we revealed, including (1) priors on $\boldsymbol{\tau}$ that are more uniform over divergence models (rather than classes of models), (2) more flexible priors on θ and τ parameters (e.g., gamma or log-normal distributions) that might increase the relative marginal likelihoods of models with more divergence times, and (3) adding more loci. The degree to which our results are contingent upon the number of taxon-pairs and sample sizes simulated here also needs to be explored. However, our results clearly demonstrate the need for (1) power analyses to accompany any inference of clustered divergences using `msBayes` and (2) analyses exploring the sensitivity of the results to prior distributions.

1.5 Conclusions

The hierarchical ABC model implemented in `msBayes` provides an appealing method of inferring the effects of historical events on diversification. Our goal was to use the model to test whether Pleistocene climate cycles and associated sea-level oscillations caused diversification across the islands of the Philippine Archipelago. Despite strong support for the recent simultaneous divergence of 22 pairs of populations across the Philippines, our simulation-based power analyses demonstrate that we were likely to infer such results. Our simulations show that `msBayes` will infer highly clustered divergence times when populations diverged randomly over the past $12N_C$ generations (7.5 million generations if $\mu = 2 \times 10^{-8}$), and will often ($> 5\%$ of replicates) infer the extreme case of one divergence event with high posterior probability (> 0.95) when divergences were random over the past $3.2N_C$ generations (2 mil-

lion generations). For our empirical system, this lack of power precludes us from ascribing biological processes to the results we obtained from `msBayes`. We also show that results of `msBayes` can be sensitive to prior distributions placed on parameters. We suggest any results of `msBayes` that are not shown to be robust to prior settings should be treated with caution. Furthermore, simulation-based power analyses should be used in cases in which clustered divergences are inferred. These analyses can provide useful guides for the range of divergence times which could have occurred while still being judged to be "simultaneous".

1.6 Tables

Table 1.1. Summary of the notation used throughout Chapter 1.

Symbol	Description
Y	Number of population pairs.
k_i	Number of loci sampled from population pair i.
K	Total number of unique loci sampled.
$X_{i,j}$	Sequence alignment of locus j sampled from population pair i.
$S^*_{i,j}$	Population genetic summary statistics calculated from $X_{i,j}$.
$\mathbf{X}$	Vector containing the sequence alignments of each locus from each population pair: $\{X_{1,1}, \ldots, X_{Y,k_Y}\}$.
$\mathbf{S}^*$	Vector containing the summary statistics of each locus from each population pair: $\{S^*_{1,1}, \ldots, S^*_{Y,k_Y}\}$.
$B_\epsilon(\mathbf{S}^*)$	Multi-dimensional Euclidean space around the observed summary statistics, $\mathbf{S}^*$.
ϵ	Radius of $B_\epsilon(\mathbf{S}^*)$, i.e., the tolerance of the ABC estimation.
$G_{i,j}$	Gene tree of the sequences in $X_{i,j}$.
$\mathbf{G}$	Vector containing the gene trees of each locus from each population pair: $\{G_{1,1}, \ldots, G_{Y,k_Y}\}$.
Ψ	Number of unique population divergence times among the Y population pairs.
T	One of the Ψ unique divergence times in $4N_C$ generations.
$\mathbf{T}$	Vector of Ψ unique divergence times: $\{T_1, \ldots, T_\Psi\}$.
τ_i	Time of divergence in $4N_C$ generations between the populations of pair i.
$t_{i,j}$	Scaled time of divergence between the populations of pair i for locus j.
$\boldsymbol{\tau}$	Vector containing the divergence times for each population pair: $\{\tau_1, \ldots, \tau_Y\}$.
$\mathbf{t}$	Vector containing the scaled divergence times of each locus from each population pair: $\{t_{1,1}, \ldots, t_{Y,k_Y}\}$.
$\theta_{D1,i}$, $\theta_{D2,i}$	θ of the 1^{st} and 2^{nd} descendent population, respectively, of pair i.
$\theta_{A,i}$	θ of the population ancestral to pair i.
$\boldsymbol{\theta_{D1}}$, $\boldsymbol{\theta_{D2}}$	Vectors $\{\theta_{D1,1}, \ldots, \theta_{D1,Y}\}$ and $\{\theta_{D2,1}, \ldots, \theta_{D2,Y}\}$, respectively.
$\boldsymbol{\theta_A}$	Vector containing the θ_A parameters for each population pair: $\{\theta_{A,1}, \ldots, \theta_{A,Y}\}$.
υ_j	θ-scaling parameter for locus j.
$\boldsymbol{\upsilon}$	Vector containing the θ-scaling parameters for each locus: $\{\upsilon_1, \ldots, \upsilon_K\}$.
α	Hyper-parameter for the shape of the gamma prior distribution on each υ.
$\zeta_{D1,i}$, $\zeta_{D2,i}$	θ-scaling parameters that determine the magnitude of the population bottleneck in the 1^{st} and 2^{nd} descendant population of pair i, respectively. The bottleneck in each descendant population begins immediately after divergence.
$\boldsymbol{\zeta_{D1}}$, $\boldsymbol{\zeta_{D2}}$	Vectors $\{\zeta_{D1,1}, \ldots, \zeta_{D1,Y}\}$ and $\{\zeta_{D2,1}, \ldots, \zeta_{D2,Y}\}$, respectively.
$\tau_{B,i}$	Proportion of time between present and τ_i when the bottleneck ends for the descendant populations of pair i.
$\boldsymbol{\tau_B}$	Vector containing the τ_B parameters for each population pair: $\{\tau_{B,1}, \ldots, \tau_{B,Y}\}$.
m_i	Symmetric migration rate between the descendant populations of pair i.
$\mathbf{m}$	Vector containing the migration rates for each population pair: $\{m_1, \ldots, m_Y\}$.
$\rho_{i,j}$	θ-scaling constant provided by the user for locus j of pair i. This constant is intended to allow the user to scale θ for differences in ploidy among loci or differences in generation times among taxa.
$\nu_{i,j}$	θ-scaling constant provided by the user for locus j of pair i. This constant is intended to allow the user to scale θ for differences in mutation rates among loci or among taxa.
$\boldsymbol{\rho}$	Vector of ploidy and/or generation time scaling constants: $\{\rho_{1,1}, \ldots, \rho_{Y,k_Y}\}$.
$\boldsymbol{\nu}$	Vector of mutation-rate scaling constants: $\{\nu_{1,1}, \ldots, \nu_{Y,k_Y}\}$.
$E(\tau)$	Mean of $\boldsymbol{\tau}$.
$Var(\tau)$	Variance of $\boldsymbol{\tau}$.
Ω	Dispersion index of $\boldsymbol{\tau}$ ($Var(\tau)/E(\tau)$).
Θ	Vector containing all the parameters of the model implemented in `msBayes`.
$f(\Theta)$	Joint prior probability distribution of the `msBayes` model.
n	Number of samples from the joint prior, $f(\Theta)$.
Λ	Vector containing the following summary of Θ: $\{\Psi, E(\tau), Var(\tau), \Omega\}$.
$\mathbf{S}$	Vector containing the summary statistics calculated from data simulated under Θ.
$\mathcal{P}_{f(\Theta)}$	Random sample of $\Lambda_1, \ldots, \Lambda_n$ from $f(\Theta)$.
$\mathbb{S}$	Summary statistic vectors $\mathbf{S_1}, \ldots, \mathbf{S_n}$ for each $\Lambda_1, \ldots, \Lambda_n$ within $\mathcal{P}_{f(\Theta)}$.
$\mathcal{P}_\epsilon$	Samples retained from $\mathcal{P}_{f(\Theta)}$ after rejection sampling. I.e., the Λ's from $\mathcal{P}_{f(\Theta)}$ with $\mathbf{S}$'s that fall within $B_\epsilon(\mathbf{S}^*)$.
$\mathcal{P}_{f(\Theta\|B_\epsilon(\mathbf{S}^*))}$	The estimate of the approximate posterior, $f(\Theta\|B_\epsilon(\mathbf{S}^*))$. I.e., the regression-adjusted $\mathcal{P}_\epsilon$.

Table 1.2. Summaries of the posterior estimates from all `msBayes` analyses run with prior settings of $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$. Mode estimates of the dispersion index (Ω) and mean ($E(\tau)$) of divergence time vector, $\boldsymbol{\tau}$, and the number of unique divergence times (Ψ) are provided, followed in square brackets by the 95% highest posterior density for ABC_{GLM} analyses and the 2.5% and 97.5% quantiles for the ABC_{LLR} analyses. The estimated posterior probability for the one divergence model ($p(\Psi = 1|B_\epsilon(\mathbf{S}^*))$) is given, followed in parentheses by the Bayes factor of the $\Psi = 1$ model compared to all other models. Ω and $E(\tau)$ are in units of N_C generations.

Prior sample	Analysis	$\hat{\Omega}$	$\hat{E(\tau)}$	$\hat{\Psi}$	$p(\Psi = 1\|B_\epsilon(\mathbf{S}^*))$
$\mathbb{S}_{stats,N=1\times10^7}$	ABC_{LLR}	0.0 [0.0–1.84×10^{-4}]	0.064 [0.043–0.092]	1.0 [1.00–1.00]	1.000 (∞)[1]
$\mathbb{S}_{stats,N=1\times10^7}$	ABC_{GLM}	1.29×10^{-4} [-1.08×10^{-6}–1.18×10^{-3}]	0.063 [0.039–0.092]	1.0 [1.00–1.06]	1.000 (∞)[1]
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	-8.36×10^{-17} [-5.49×10^{-6}–5.89×10^{-3}]	0.062 [0.035–0.092]	1.0 [1.00–1.06]	1.000 (∞)[1]
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	6.53×10^{-4} [-2.18×10^{-5}–1.89×10^{-3}]	0.071 [0.042–0.103]	1.0 [1.00–1.06]	0.999 (2.1×10^4)
$\mathbb{S}_{PLS,N=1\times10^7}$	ABC_{LLR}	0.0 [0.0–6.38×10^{-4}]	0.084 [0.036–0.136]	1.0 [1.00–1.00]	0.993 (2979)
$\mathbb{S}_{PLS,N=1\times10^7}$	ABC_{GLM}	-9.80×10^{-17} [-7.56×10^{-6}–7.59×10^{-3}]	0.082 [0.024–0.127]	1.0 [1.00–1.14]	0.977 (892)
$\mathbb{S}_{PLS,N=5\times10^6}$	ABC_{GLM}	-9.09×10^{-17} [-1.11×10^{-5}–1.31×10^{-2}]	0.064 [0.013–0.126]	1.0 [1.00–1.13]	0.992 (2604)
$\mathbb{S}_{PLS,N=5\times10^6}$	ABC_{GLM}	-9.80×10^{-17} [-7.64×10^{-6}–8.26×10^{-3}]	0.086 [0.029–0.143]	1.0 [1.00–1.08]	0.966 (597)
13 Mindanao population-pairs					
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{LLR}	2.56×10^{-3} [0.0–6.47×10^{-3}]	0.079 [0.034–0.122]	1.0 [1.00–1.00]	0.987 (1594)
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	-8.03×10^{-17} [-6.00×10^{-6}–1.36×10^{-2}]	0.070 [0.027–0.115]	1.0 [1.00–1.08]	0.962 (532)
9 Negros-Panay population-pairs					
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{LLR}	0.0 [0.0–1.08×10^{-2}]	0.060 [0.012–0.099]	1.0 [1.00–1.00]	1.000 (∞)[1]
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	-5.96×10^{-17} [-5.14×10^{-5}–4.53×10^{-2}]	0.055 [0.014–0.095]	1.0 [1.00–1.12]	0.999 (2.1×10^4)

[1] An estimate of 1.0 for a posterior probability is an artifact of sampling error.

Table 1.3. Posterior estimates of `msBayes` from the empirical data under three different model priors. Mode estimates of the dispersion index (Ω) and mean ($E(\tau)$) of divergence time vector, $\boldsymbol{\tau}$, and the number of unique divergence times (Ψ) are provided, followed in square brackets by the 95% highest posterior density for ABC_{GLM} analyses and the 2.5% and 97.5% quantiles for the ABC_{LLR} analyses. The estimated posterior probability for the inferred divergence model ($p(\Psi = \hat{\Psi}|B_\epsilon(\mathbf{S}^*))$) is given, followed in parentheses by the Bayes factor of the inferred model compared to all other models. The estimated posterior probability that Ψ is not one ($p(\Psi \neq 1|B_\epsilon(\mathbf{S}^*))$) is also given. Ω and $E(\tau)$ are in units of N_C generations.

Prior sample	Analysis	$\hat{\Omega}$	$\hat{E(\tau)}$	$\hat{\Psi}$	$p(\Psi = \hat{\Psi}\|B_\epsilon(\mathbf{S}^*))$	$p(\Psi \neq 1\|B_\epsilon(\mathbf{S}^*))$
$f(\Theta)$: $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$						
$\mathbb{S}_{stats,N=1\times10^7}$	ABC_{LLR}	0.0 [0.0–1.84$\times10^{-4}$]	0.064 [0.043–0.092]	1.00 [1.00–1.00]	0.992 (2604)	0.008
$\mathbb{S}_{stats,N=1\times10^7}$	ABC_{GLM}	1.29$\times10^{-4}$ [-1.08$\times10^{-6}$–1.18$\times10^{-3}$]	0.063 [0.039–0.092]	1.00 [1.00–1.06]	1.000 (∞)[1]	0.000
$f(\Theta)$: $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$						
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{LLR}	0.217 [0.112–0.381]	0.391 [0.282–0.485]	2.00 [2.00–2.00]	0.986 (1479)	0.986
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	0.249 [0.030–0.475]	0.399 [0.271–0.513]	2.01 [1.39–3.40]	0.785 (77)	0.970
$f(\Theta)$: $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$						
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{LLR}	0.421 [0.200–0.639]	0.465 [0.336–0.566]	2.00 [2.00–4.00]	0.784 (76)	0.999
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	0.441 [0.190–0.711]	0.513 [0.358–0.601]	2.04 [1.51–5.16]	0.343 (11)	0.986

[1] An estimate of 1.0 for a posterior probability is an artifact of sampling error.

1.7 Figures

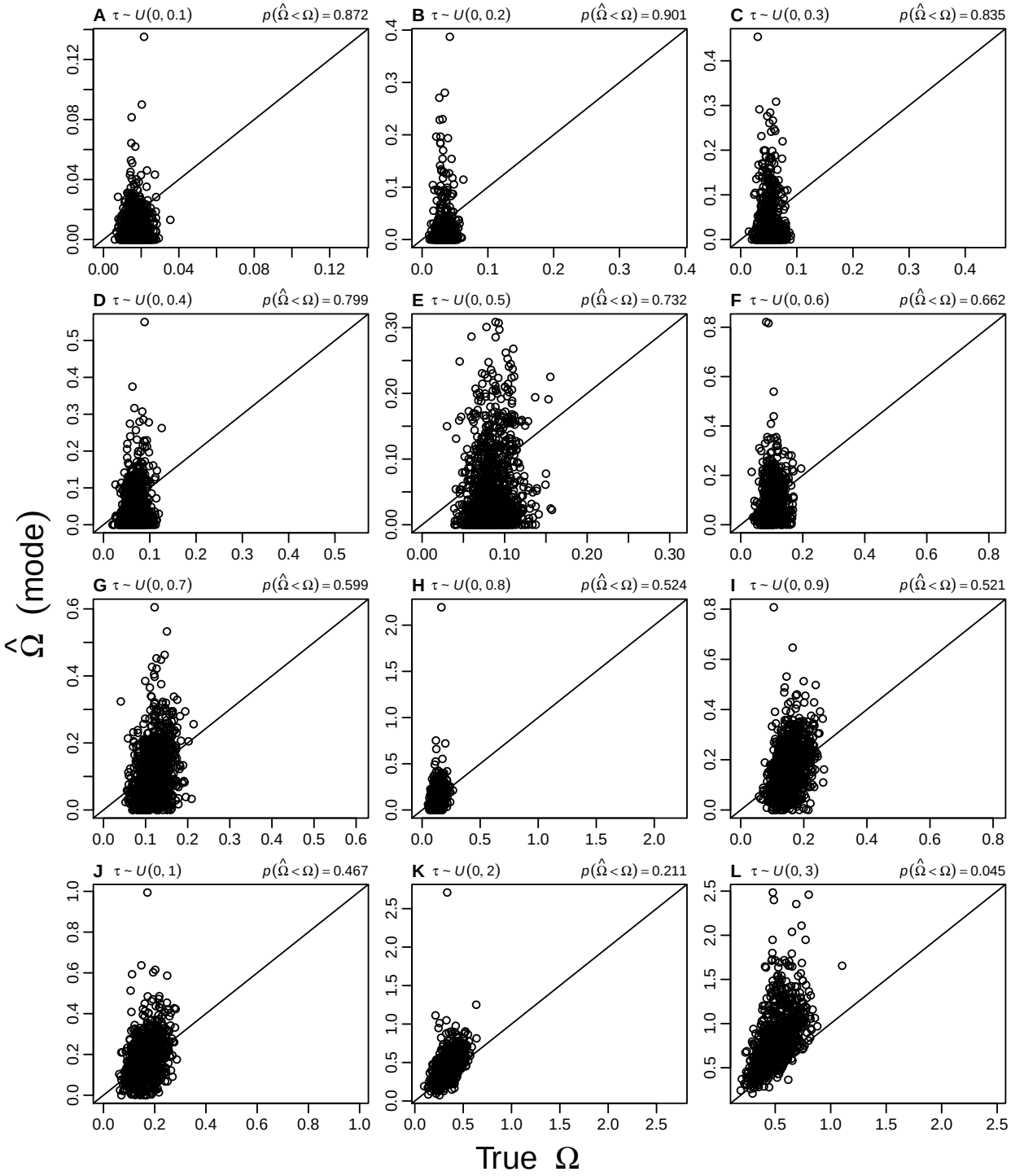

Figure 1.1. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

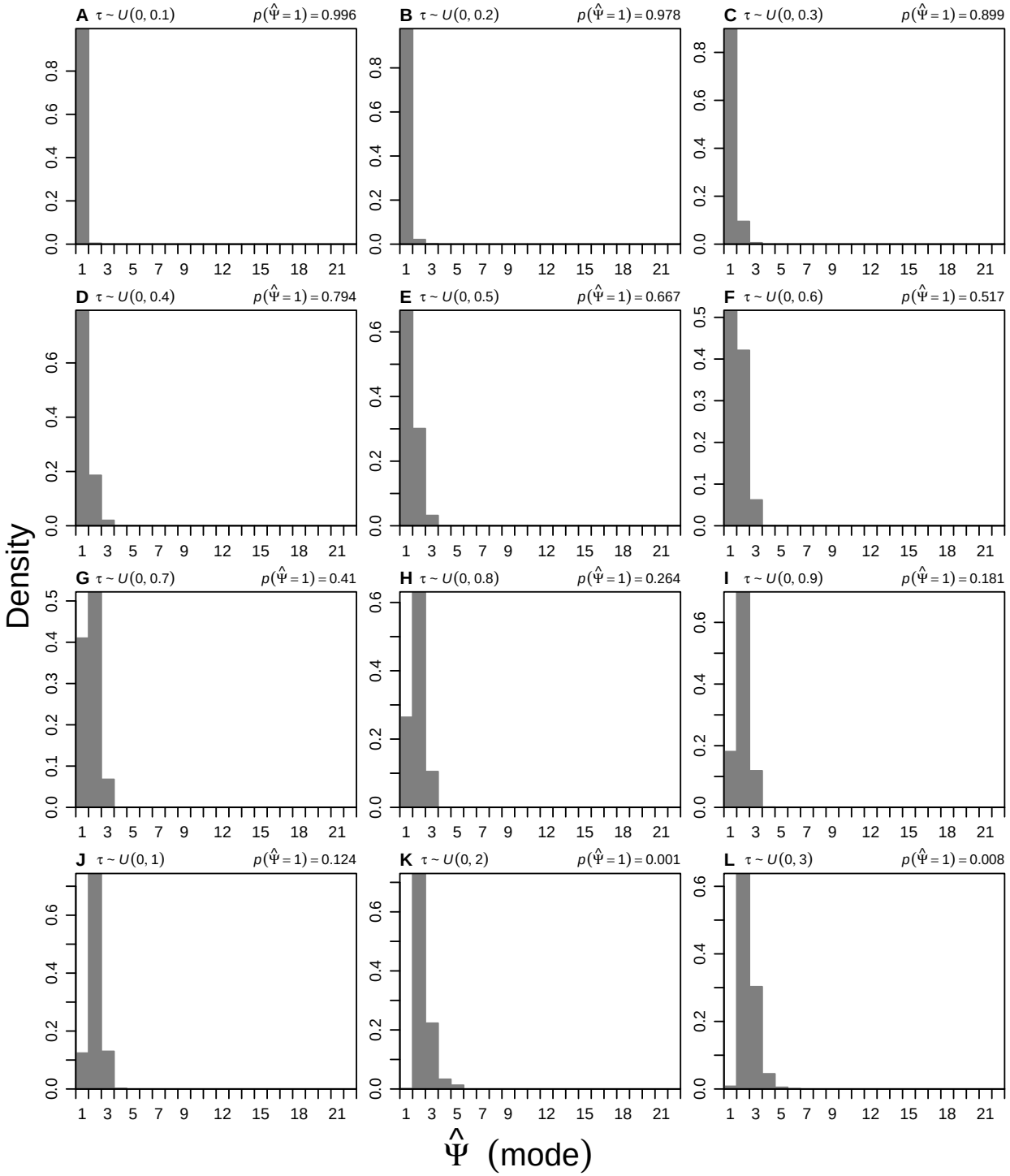

Figure 1.2. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

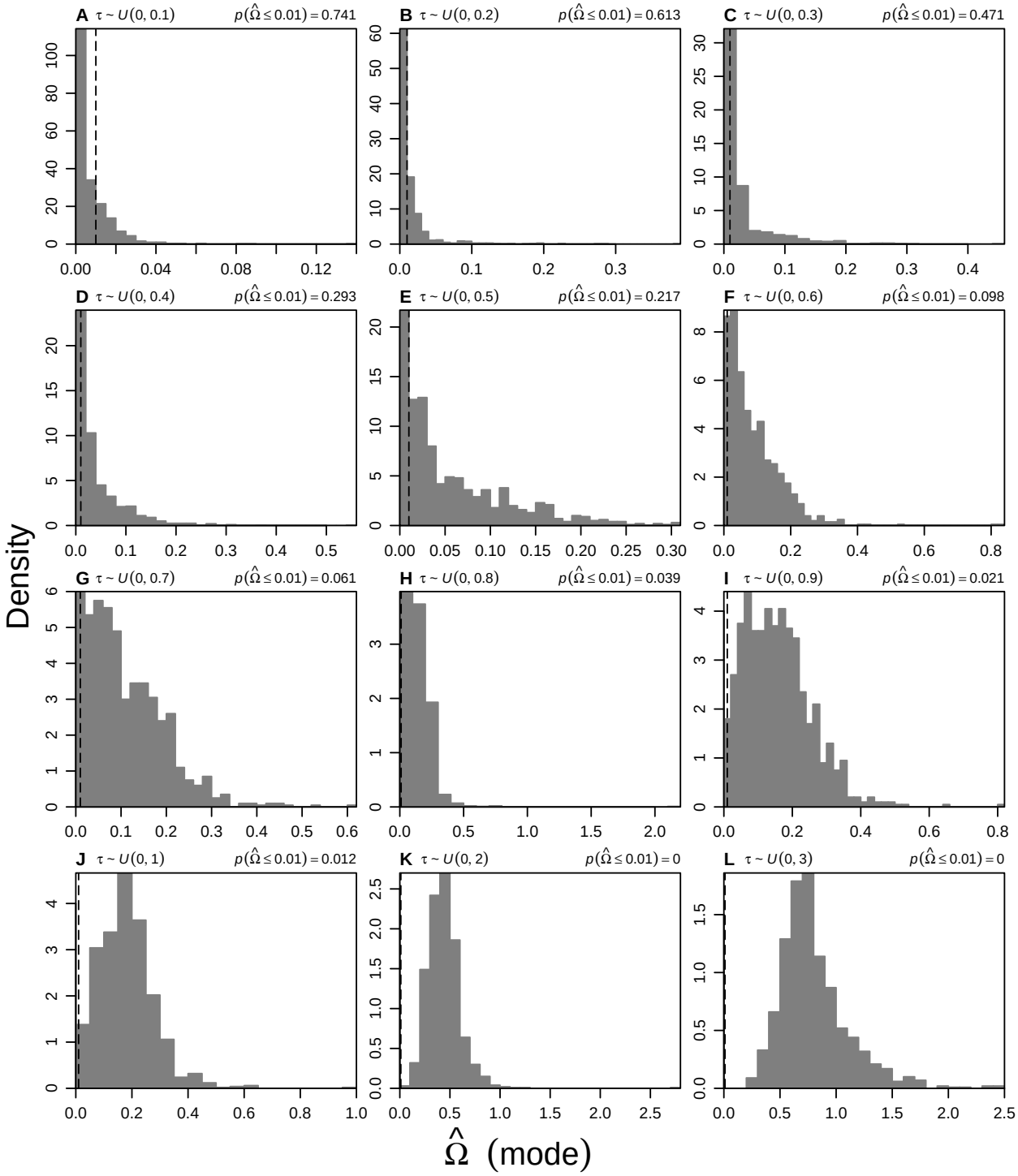

Figure 1.3. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

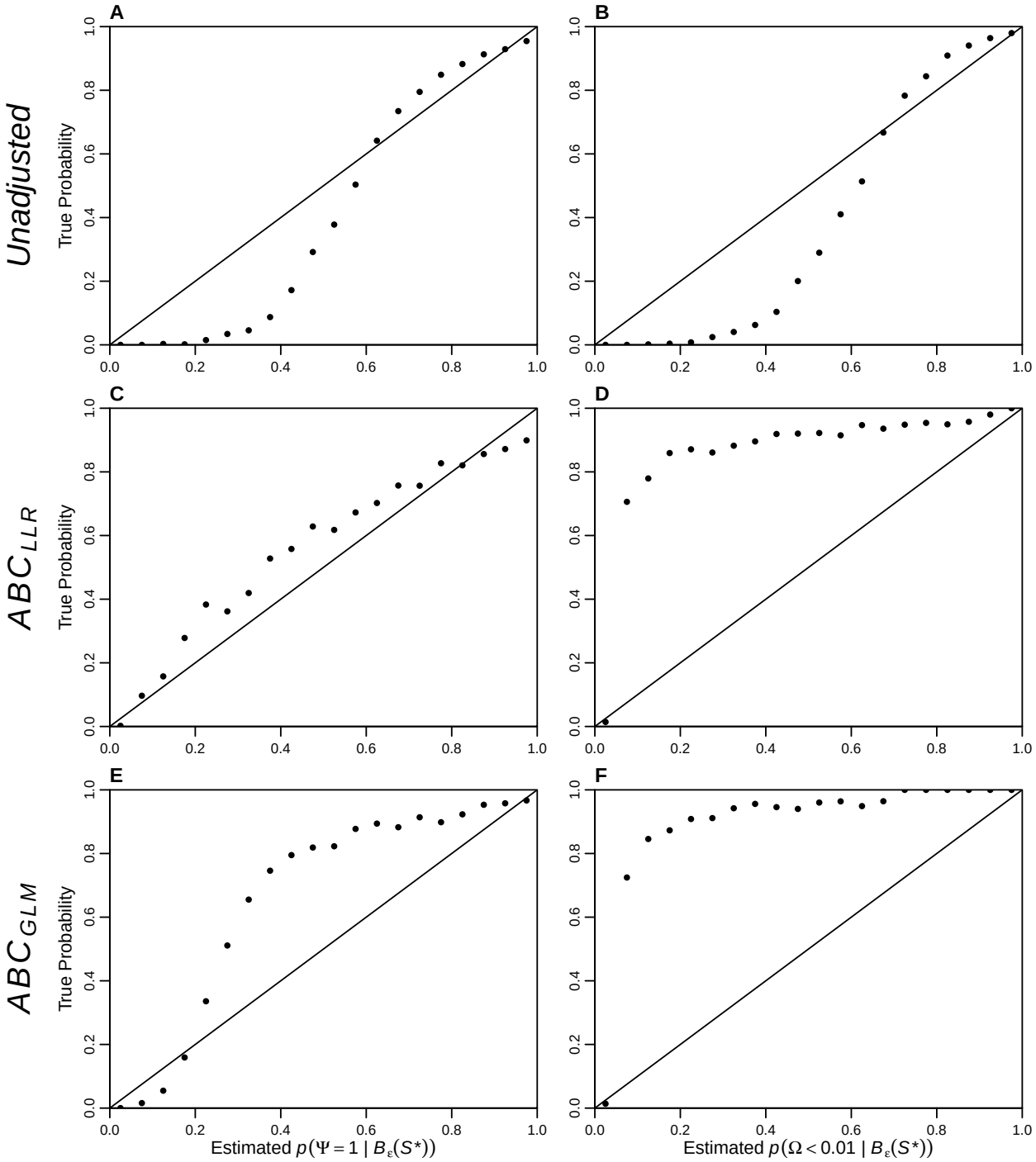

Figure 1.4. The relationship between the posterior and true probability of (A, C & E) $\Psi = 1$ and (B, D & F) $\Omega < 0.01$ based on 100,000 simulations. The results are based on the (A & B) unadjusted ($\mathcal{P}_\epsilon$), (C & D) ABC_{LLR}-adjusted, and (E & F) ABC_{GLM}-adjusted posterior estimate from each simulation replicate. All simulated replicates were generated under the model prior (i.e., the ideal situation where the prior model is correct). Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0001, 0.05)$, and $\theta_A \sim U(0.0001, 0.025)$, and the number of samples from the prior was 2×10^6. The simulated data structure was 10 population pairs, with a single 1000 bp locus sampled from 10 individuals from each population. The 100,000 estimates of the posterior probability of one divergence event were assigned to 20 bins of width 0.05. The estimated posterior probability of each bin is plotted against the proportion of replicates in that bin with a true value consistent with one divergence event (i.e., $\Psi = 1$ or $\Omega < 0.01$).

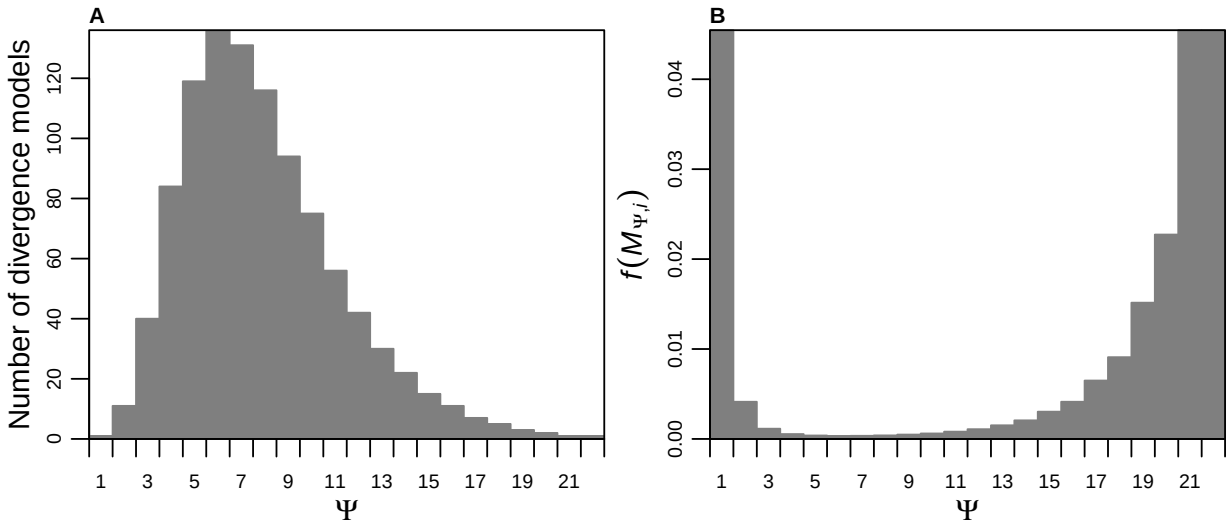

Figure 1.5. The (A) number of unique divergence models across all possible numbers of divergence time categories (Ψ) when $Y = 22$, and (B) the prior probability of any one of the unique divergence time models within each Ψ.

Chapter 2

Why you should not fix a biased model-choice method by adding an additional dimension of model choice: A reply to Hickerson et al.

2.1 Introduction

Recently, the journal *Evolution* has served as a medium for a discussion of the potential pitfalls of approximate Bayesian methods of comparative phylogeographical model choice and prior selection. The discussion has centered around the approximate Bayesian model choice method, `msBayes`, which estimates the temporal distribution of divergences among co-distributed pairs of taxa (Huang et al., 2011) by sampling over all unordered models of divergence. Oaks et al. (2013) published results of simulation-based power analyses showing that `msBayes` can often be biased toward inferring models of temporally clustered divergence times across taxon pairs. Hickerson et al. (2013) published a response to Oaks et al. (2013) where they present an empirical Bayes model-averaging approach that they conclude cir-

cumvents the poor behavior of the method revealed by Oaks et al. (2013). These papers are largely in agreement, with Hickerson et al. (2013) reiterating much of the discussion of Oaks et al. (2013) regarding the impact of broad uniform priors on Bayesian model choice. The main difference in perspectives are centered around (1) the mechanism by which broad uniform priors cause the poor behavior of `msBayes`, and (2) how to potentially mitigate this issue.

Oaks et al. (2013) suggest the primary cause of the poor behavior of the method is likely the low marginal likelihoods of parameter-rich models integrated over vast parameter space with low probability of producing the data, yet relatively high prior density (Lindley, 1957). Note, this suggests the bias is extrinsic to `msBayes`, and the numerical approximation machinery of `msBayes` could be sound. Hickerson et al. (2013) suggest the bias is intrinsic to `msBayes`, concluding the method's rejection algorithm is inefficient and will be increasingly biased as the overall space of the model increases, either as a function of the number of taxon pairs or the width of the uniform priors on nuisance parameters. Hickerson et al. (2013) support their position by giving a probabilistic argument that focuses on only one of the three prior models used by Oaks et al. (2013). We show that this argument is based on an empirical Bayesian reasoning and questionable assumptions, and does not explain the bias of the method found in many of the analyses of Oaks et al. (2013). Furthermore, we present results of additional analyses that, when combined with the results of Oaks et al. (2013), strongly suggest small marginal likelihoods of models with more parameter space is a primary mechanism by which uniform priors are causing the bias of `msBayes`.

To mitigate the bias, Oaks et al. (2013) suggest alternative prior probability distributions on divergence models and nuisance parameters could increase the likelihoods of models with more divergence time parameters. They also describe simulation-based procedures for determining power, accuracy, and robustness of the method given a dataset, and recommend any application of `msBayes` should be accompanied by such procedures. Alternatively, Hickerson et al. (2013) present an approximate Bayesian model-averaging approach for accommodat-

ing uncertainty in selecting among empirically guided priors, and champion the method as a means of avoiding the pitfalls raised by Oaks et al. (2013). We discuss errors in this approach that render some of their results invalid and leave the remaining results difficult to interpret. Furthermore, we follow the recommendations of Oaks et al. (2013) and present simulation-based assessments of Hickerson et al.'s (2013) approach. Our results demonstrate that the empirical Bayesian approach of Hickerson et al. (2013) is biased toward clustered divergence models and predisposed to excluding true parameter space. The model-averaging approach provides another means by which the model can "escape" large parameter space, and the method's bias toward smaller models remains. This manifestation of the bias often excludes the true parameter space due to the use of narrow uniform priors. Furthermore we discuss the potential theoretical and practical problems of empirical Bayesian model choice.

2.2 An error in Hickerson et al.'s re-analysis of the Philippines data

Hickerson et al. (2013) re-analyzed the dataset of Oaks et al. (2013) using a model-averaging approach, where they placed a discrete uniform prior over eight different prior models (see Table 1 of Hickerson et al. (2013)). However, there is a fundamental error in the methodology of Hickerson et al. (2013): their model mixes different time units.

Each of the eight prior models used in the re-analysis by Hickerson et al. (2013) has one of two priors on the mean size of the descendant populations of each taxon pair: $\theta_D \sim U(0.0001, 0.1)$ or $\theta_D \sim U(0.0005, 0.04)$. As described in Oaks et al. (2013), the divergence time parameters of the model implemented in `msBayes` are scaled relative to a constant reference population size, θ_C. This reference population size is defined in terms of the upper limit of the uniform prior on the mean size of the descendant populations, θ_D, such that for the prior $\theta_D \sim U(a_{\theta_D}, b_{\theta_D})$, $\theta_C = b_{\theta_D}/2$. Thus, the model used by Hickerson et al. (2013) mixes two different units of time. In other words, some of their prior and posterior samples

are in units of $0.05/\mu$ generations, whereas others are in units of $0.02/\mu$ generations.

The fact that their posterior samples are in different units makes the results of Hickerson et al. (2013) difficult to interpret, and renders their regression-adjusted results invalid. A fundamental assumption of regression is that all of the values of the response variable are in the same units. Thus, the results in sections "Using ABC Model Comparison to Weight Alternative Priors for the Philippine Vertebrate Data" and "Improved Sampling Efficiency by Prior Weighting Supports Asynchronous and Recent Divergence for the Philippines Vertebrate Data" and presented in Figure 2 of Hickerson et al. (2013) should be disregarded. The error is easily illustrated by re-plotting their results with the different time units indicated (Figure 2.1).

Another important characteristic of the approach used by Hickerson et al. (2013) that makes their results difficult to interpret is their use of narrow, empirically guided uniform priors in an approximate-Bayesian model-choice framework. We discuss the risks associated with this approach in the next section, and show how the results of Hickerson et al. (2013) likely exclude true values of the model's parameters.

2.3 The potential implications of empirical Bayesian model choice

Hickerson et al. (2013) repeatedly refer to the prior distributions used by Oaks et al. (2013) as "poorly selected." However, this is misleading, as Oaks et al. (2013) discuss in detail how they selected their priors (see their section "Specifying and simulating the joint prior"). They mention how the use of uniform prior distributions in `msBayes` for many of the model's parameters necessitates awkward broad priors to avoid excluding the truth, *a priori*. What Hickerson et al. (2013) mean by "poorly selected" is that Oaks et al. (2013) did not use their data to inform their initial prior distributions (i.e., they did not use an empirical Bayesian approach). Oaks et al. (2013) did use empirically informed priors when assessing

both the prior sensitivity of the method and determining the power of `msBayes` under real-world conditions with highly informative priors. Oaks et al. (2013) discuss the potential dangers of taking an empirical Bayesian approach to model choice (see the last paragraph of "Assessing prior sensitivity of `msBayes`" in Oaks et al. (2013)), and we expand on this here by exploring both the theoretical and practical implications of using empirically informed priors for Bayesian model choice.

2.3.1 Theoretical implications of empirical priors for Bayesian model choice

Bayesian inference is a method of inductive learning in which Bayes' rule is used to update our beliefs about a model as new information becomes available. If we let $\mathbf{\Theta}$ represent the set of all possible parameter values, we can define a prior distribution for all $\theta \in \mathbf{\Theta}$ such that $p(\theta)$ describes our belief that any given θ is the true value of the parameter. If we let $\mathcal{X}$ represent all possible datasets then we can define a sampling model for all $\theta \in \mathbf{\Theta}$ and $X \in \mathcal{X}$ such that $p(X|\theta)$ measures our belief that any dataset X will be generated by any model state θ. After collecting a new dataset X_i, we can use Bayes' rule to calculate the posterior distribution

$$p(\theta \,|\, X_i) = \frac{p(X_i \,|\, \theta)p(\theta)}{p(X_i)}, \tag{2.1}$$

where

$$p(X_i) = \int_\theta p(X_i \,|\, \theta)p(\theta)d\theta. \tag{2.2}$$

The posterior distribution is a measure of our beliefs after seeing the new information.

This is an elegant method of updating our beliefs as data are accumulated. However, this all hinges on the fact that the prior ($p(\theta)$) is defined for all possible parameter values

independently of the new data being analyzed. Any other datasets or external information can safely be used to inform our beliefs about $p(\theta)$. However, if the same data are used to both inform the prior and calculate the posterior, the prior becomes conditional on the data, and Bayes' rule breaks down.

This is not to say that empirical Bayesian approaches are not useful. Empirical Bayes is a well studied branch of Bayesian statistics that has given rise to many powerful inference methods. However, empirical Bayes methods have an uncertain theoretical basis and do not yield a valid posterior distribution from Bayes' rule (e.g., empirical Bayesian estimates of the posterior are often too narrow, off-center, and incorrectly shaped; Morris, 1983; Laird and Louis, 1987; Carlin and Gelfand, 1990; Efron, 2013). Nevertheless, empirical Bayes methods can provide a powerful means of parameter estimation that often exhibit favorable frequentist properties. Furthermore, many post-hoc correction methods have been developed for estimating confidence-intervals from empirical Bayes estimates of posterior distributions that often exhibit well-behaved frequentist coverage probabilities (Morris, 1983; Laird and Louis, 1987, 1989; Carlin and Gelfand, 1990; Hwang et al., 2009).

Whereas empirical Bayes approaches can provide powerful methods for parameter estimation, a theoretical justification for empirical Bayes approaches to model choice is questionable. In Bayesian model choice, the goal is not to estimate parameters, but to estimate the relative probabilities of candidate models. Unlike parameter estimates, posterior probabilities represent a summary of the entire posterior distribution. Thus, given that empirical Bayesian posterior distributions are not accurate, there is no guarantee regarding the accuracy of probabilities summarized from them. In addition to these theoretical concerns there are practical dangers to using narrow, empirical, uniform priors for a method that has already been shown to be biased toward models with less parameter space.

2.3.2 Practical concerns about empirically informed uniform priors for Bayesian model choice

The results of Hickerson et al.'s 2013 reanalysis of the Philippine dataset strongly favored models with the narrowest, empirically informed prior on divergence time, and thus their model-averaged posterior estimates are dominated by models M_1 and M_2 (see Table 1 of Hickerson et al. (2013)). This is concerning, because the narrowest τ prior used by Hickerson et al. (2013) ($\tau \sim U(0, 0.1)$) likely excludes the true divergence times for at least some of the Philippine taxa, a major problem when using uniform priors. Hickerson et al. (2013) set this prior to match the 95% highest posterior density (HPD) interval for the mean divergence time estimated under one of the priors used by Oaks et al. (2013) (see Tables 2 and 3 of Oaks et al. (2013)). Given this interval estimate is for the *mean* divergence time across all 22 taxa, it may be inappropriate to set this as the limit on the prior, because many of the taxon pairs are expected to have diverged at times older than the upper limit. Furthermore, this prior is *excluded* from the 95% HPD interval estimates of the mean divergence time under the other two priors explored by Oaks et al. (2013) (under these priors the 95% HPD is approximately 0.3–0.6; see Table 6 of Oaks et al. (2013)). Thus, the results of Hickerson et al. (2013) indicate that integration over parameter space with low likelihood is biasing the method toward models with smaller parameter space, and as a consequence is biasing the method toward models that exclude the truth (i.e., toward estimating model-averaged posteriors dominated by models that exclude true values of parameters of the model).

We explored the propensity of the model-averaging approach of Hickerson et al. (2013) to exclude the truth in two ways. First, we re-analyzed the Philippines dataset using the model-averaging approach of Hickerson et al. (2013), but set one of prior models with a uniform prior on divergence times that is unrealistically narrow, and almost certainly excludes most, if not all, of the true divergence times of the 22 taxon pairs. If small likelihoods of large models cause the method to prefer models with less parameter space, we expect `msBayes` will preferentially sample from this incorrect model yielding a marginal posterior that is

incorrect (i.e., the model-averaged posterior will be dominated by an incorrect model that excludes the truth). Second, we generated simulated datasets for which the divergence times are drawn from an exponential distribution and applied the approach of Hickerson et al. (2013) to each of them to see how often the method excludes the truth.

For our re-analysis of the Philippines dataset we used the model-averaging approach of Hickerson et al. (2013), but with a reduced set of prior models to avoid mixing time units. We used five prior models, all of which had priors on population sizes of $\theta_D \sim U(0.0001, 0.1)$ and $\theta_A \sim U(0.0001, 0.05)$. Following Hickerson et al. (2013), each of these models had the following priors on divergence time parameters: M_1, $\tau \sim U(0, 0.1)$; M_2, $\tau \sim U(0, 1)$; M_3, $\tau \sim U(0, 5)$; M_4, $\tau \sim U(0, 10)$; and M_5, $\tau \sim U(0, 20)$. We simulated 1×10^6 random samples from each of the models for a total of 5×10^6 prior samples. For each model, we retained the 10,000 samples with the smallest Euclidean distance from the observed summary statistics, standardizing the statistics using the prior means and standard deviations of the given model. From the remaining 50,000 samples, we then retained the 10,000 samples with the smallest Euclidean distance from the observed summary statistics, this time standardizing the statistics using the prior means and standard deviations across all five models. We then repeated this analysis twice, replacing the M_1 with M_{1A} and M_{1B}, which differ only by having priors on divergence times of $\tau \sim U(0, 0.01)$ and $\tau \sim U(0, 0.001)$, respectively. While we suspect the prior of $\tau \sim U(0, 0.1)$ used by Hickerson et al. (2013) likely excludes the true divergence times of at least some of the 22 taxa, we are nearly certain that these narrower priors are incorrect and exclude most, if not all, of the divergence times of the Philippine taxa.

Our results show that the model-averaging approach of Hickerson et al. (2013) does not reduce the method's bias toward models with less parameter space, but rather allows it to manifest in a rather dangerous way. The method strongly prefers the prior model with the narrowest distribution on divergence times across all three of our analyses, even when this model is almost certainly incorrect and excludes the true divergence times of the

Philippine taxa (Table 2.1). Unfortunately, "checking" the priors by plotting the summary statistics from 1000 random samples of each prior model along the first two orthogonal axes of a principle component analysis, as recommended by Hickerson et al. (2013), provides no warning of a problem (Figure 2.2). Given that the results of Hickerson et al. (2013) strongly prefer the models with the narrowest prior on divergence times, it seems quite likely that their model-averaged results are dominated by models that exclude at least some of the 22 true divergence times, making their results difficult to interpret (in addition to the error of time units).

To better quantify the propensity of Hickerson et al.'s (2013) approach to exclude the truth, we simulated 1000 datasets in which the divergence times for the 22 population pairs are drawn randomly from an exponential distribution with a mean of 0.5 ($\tau \sim Exp(2)$). All other parameters were identically distributed as the M_1–M_5 models (Table 2.1). We then repeated the analysis described above using 1×10^6 random samples from prior models M_1, M_2, M_3, M_4, and M_5, retaining 1000 posterior samples for each of the 1000 simulated datasets.

For each simulation replicate, we estimated the Bayes factor in favor of excluding the truth as the ratio of the posterior to prior odds of excluding the true value of at least one parameter. Whenever the Bayes factor preferred a model excluding the truth we counted the number of 22 true divergence times that were excluded by the preferred model. Our results show that the model-averaging approach of Hickerson et al. (2013) excludes the true values of parameters in 97% of the replicates (90% with GLM-regression adjustment), excluding up to 21 of the 22 true divergence times (Figure 2.3). We also used the mode estimate of the preferred model for each replicate to estimate the number of true values excluded, which produced very similar results. Furthermore, the posterior probability of excluding at least one true parameter value is very high in nearly all of the replicates (Figure 2.4). Using a Bayes factor of greater than 10 as a criterion for strong support, 66% of the replicates (87% with GLM-regression adjustment) strongly support the exclusion of true values (Figure 2.4).

The results of our empirical and simulation-based analyses clearly demonstrate the danger of using narrow, empirically guided uniform priors in a Bayesian model-averaging framework. The consequence of this approach is a high risk of obtaining a model-averaged posterior estimate that is heavily weighted toward incorrect models that exclude true values of model parameters.

2.3.3 Additional thoughts on empirical priors in Bayesian model choice

Given all of the theoretical and practical issues with empirical Bayes approaches to Bayesian model choice discussed in the proceeding sections, it is quite clear why one should use caution before overly criticizing an investigator's choice of priors after having seen the resulting posterior. As discussed by Oaks et al. (2013), prior to analyzing the data, there was a large amount of uncertainty regarding the divergence times of the 22 population pairs under study. Two of these pairs represent distinct species, and the taxonomy of many groups in the Philippines has repeatedly been shown to mask deeply divergent lineages (Brown et al., 2008; Linkem et al., 2010; Siler et al., 2010; Welton et al., 2010; Siler et al., 2011b,a, 2012; Brown and Stuart, 2012; Linkem and Brown, 2013). Oaks et al. (2013) also discuss how the sole use of uniform priors in `msBayes` makes it very difficult for investigators to express their prior uncertainty without putting high prior density in regions of improbable parameter space. The alternative, as shown above, is to risk excluding the truth before performing the analysis.

While our results above clearly demonstrate the risks inherent to the empirical Bayesian model-choice approach used by Hickerson et al. (2013), one could justify such risk if the method does indeed increase the power of the method and thus decrease bias toward clustered divergence models, as claimed by Hickerson et al. (2013). In the next section, we use simulation-based analyses to evaluate the power of Hickerson et al.'s (2013) approach, which show the method is still biased toward highly clustered models of divergence.

2.4 Assessing the power of the model-averaging approach of Hickerson et al. (2013)

As recommended by Oaks et al. (2013), we perform simulation-based analyses to explore the power of the approximate Bayesian model-averaging approach proposed by Hickerson et al. (2013). Following Oaks et al. (2013), we simulated 1000 datasets with τ for each of the 22 population pairs randomly drawn from a uniform distribution, $U(0, \tau_{max})$, where τ_{max} was set to: 0.2, 0.4, 0.6, 0.8, 1.0, and 2.0, in $4N_C$ generations. All other parameters were identically distributed as the prior. We use the same 5×10^6 samples from the same set of five prior models as above (M_1, M_2, M_3, M_4, and M_5). As described above, for each simulated dataset, we approximated the posterior by retaining 1000 samples from the prior with the smallest Euclidean distance from the true summary statistics. In total, we analyzed 6000 replicate datasets, retaining 1000 model-averaged posterior samples for each of them.

We find that the approach of Hickerson et al. (2013) struggles to estimate the dispersion index of divergence times (Ω) across most of the τ_{max} we simulated, whether evaluating the unadjusted (Figure 2.5) or GLM-adjusted (Figure 2.6) posterior estimates. The method only estimates Ω relatively well when the simulated distribution of divergence times is identical to one of the prior models (Figures 2.5E & 2.6E). This is consistent with the conclusion of Oaks et al. (2013) that `msBayes` lacks robustness and is highly sensitive to the prior distribution deviating from the true, underlying distribution of the data.

Furthermore, our results demonstrate that the approach of Hickerson et al. (2013) consistently infers highly clustered divergences across all the τ_{max} we simulated (Figure 2.9). The method is most likely to infer the extreme case of a single divergence event when populations diverged randomly over the past $4N_C$ generations. Even when divergences are random over the past $8N_C$ generations, the most likely inference is only two divergence events, and a single divergence is still estimated in more than 10% of the replicates. It is very interesting to note that as τ_{max} increases, but before the estimates are finally pulled away from $\Psi = 1$,

the distribution of Ψ estimates closely mirror the odd U-shaped prior on divergence models used by `msBayes` (see Figure 2.9E and Oaks et al.'s (2013) Figure 5B). This supports the conclusions of Oaks et al. (2013) that this U-shaped prior coupled with the poor marginal likelihoods of models with many τ parameters, is a major cause of the method's bias toward clustered divergence models.

Looking at our simulation results in terms of the posterior probability of the dispersion index of divergence times supporting the extreme case of one divergence event (i.e., $p(\Omega < 0.01 \,|\, B_\epsilon(\mathbf{S}^*))$), we find the method strongly supports one divergence in greater than 27% of the replicates across all the τ_{max} we simulated (Figure 2.8). Following Hickerson et al. (2013), we use a Bayes factor of greater than 10 as the criterion for incorrect inference of a single divergence event. There is strong support for a single divergence event in more than 90% of the replicates when divergences are random over the past $2.4N_C$ generations, and more than 60% when over the past $3.2N_C$ generations or less (Figure 2.8).

Contrary to Hickerson et al.'s (2013) claim that their model-averaging approach "can discriminate complex multispecies histories and correctly reject synchronous divergence, even when discrete divergence times differ by much less than ... millions of generations," our results show that method is biased toward supporting the extreme case of a single divergence event when populations diverged randomly over the last $8N_C$ generations. To put this on the scale roughly consistent with a vertebrate mitochondrial locus, assuming a mutation rate of 2×10^{-8} per site per generation, this translates to 5 million generations. Assuming a mutation rate consistent with nuclear loci of 1×10^{-9}, this is 100 million generations.

The results of our power analyses further demonstrate the propensity of Hickerson et al.'s (2013) approach to exclude true parameter values. Across all but one of the τ_{max} we simulated, the method excludes the truth in a large proportion of replicates, and across many of the τ_{max} it will exclude a large proportion of the true divergence time values (Figure 2.9). The posterior probability of excluding at least on true divergence value is also quite high across many of the τ_{max} (Figure 2.10). Only when the data are identically distributed as one

of the prior models does the method avoid excluding the truth more than 5% of the time (Figure 2.9E). Again, this demonstrates the method's lack of robustness.

Given our results, we conclude that the approach of Hickerson et al. (2013) excludes regions of parameter space containing the truth, and lacks power to detect random variation in divergence times over a scale of $8N_C$ generations or more. This roughly translates to millions of generations for mitochondrial loci, and hundreds of millions of generations for nuclear loci.

2.5 The power analysis of Hickerson et al. (2013)

Hickerson et al. (2013) present a power analysis in which they find that `msBayes` has the power to infer multiple divergence events when divergence times are random over hundreds of thousands of generations (rather than millions as demonstrated by Oaks et al. (2013)). It is not surprising that under limited parameter space the method has increased power to detect temporal variation in divergences over narrower time windows than millions of generations. However, what is important to consider is whether those conditions are relevant to real-world applications of the method. Oaks et al. (2013) explore the behavior of the method under three divergence time priors, as narrow as 0–5 coalescent units. This is quite narrow considering that this expresses the prior belief that all 22 taxon pairs diverged within this window. Hickerson et al. (2013) limit their power analysis to a single prior of 0–1 coalescent unit. This prior setting assumes that there is enough a priori information to be 100% certain that all taxa (18 for their simulations) diverged within the last $4N_C$ generations. This does not seem applicable to most empirical systems. Furthermore, even when such extensive prior information is available, to only be able to detect multiple divergences on the scale of hundreds of thousands of generations does not seem very powerful.

Hickerson et al. (2013) present information only on the extreme case of a single divergence event. As discussed in Oaks et al. (2013), it seems odd to consider the estimation of

two divergence events as a success when divergence is random over hundreds of thousands (or millions) of generations. In an empirical system such as the Philippines, where island fragmentation has occurred at least 10 times over the past several millions of years (Haq et al., 1987; Rohling et al., 1998; Siddall et al., 2003; Miller et al., 2005), an estimate where 22 taxa share even a handful of divergence events would be of biogeographic interest.

Lastly, Hickerson et al. (2013) translate their results from units of $4N_C$ generations to generations assuming a mutation rate of 1.92×10^{-8} mutations per site per generation. If we translate their results to a scale more consistent with rates of mutation of nuclear loci (1×10^{-9}), even under the very optimistic prior settings used by Hickerson et al. (2013) the method can only reliably reject the extreme case of one divergence event when divergences are random over a period of more than 3 million generations. Again, considering Hickerson et al. (2013) are assuming prior knowledge that all population pairs diverged within the last coalescent unit, the method has low power to detect random variation in divergence times.

2.6 Model likelihoods or insufficient prior sampling as the cause of bias in `msBayes`

One of the points of disagreement between Hickerson et al. (2013) and Oaks et al. (2013) is the underlying mechanism causing the bias toward models with clustered divergences (i.e., models with few divergence events). Oaks et al. (2013) present simulation-based analyses that suggest the broad uniform prior on divergence time parameters decrease the marginal likelihoods of models with more divergence time parameters, whereas Hickerson et al. (2013) argue that the bias is inherent to the inefficient rejection algorithm implemented in `msBayes`.

Hickerson et al. (2013) present an interesting probabilistic argument to show that insufficient prior sampling is to blame for the bias. They argue the widest of the three priors used by Oaks et al. (2013) was very unlikely to produce any prior samples with large numbers of divergence times that were consistent with the Philippine data. There are a few issues

with their argument. First, the probabilities they present assume the gene divergence times estimated by Oaks et al. (2013) are correct. The sole purpose of the estimates of ultrametric gene trees presented by Oaks et al. (2013) was to provide a very rough comparison of the gene divergence times across the 22 taxa. These analyses assumed an arbitrary strict clock of 2×10^{-8} for all taxa, and are, of course, subject to estimation error. Furthermore, the branch-length units of the gene trees are in millions of years, whereas the divergence time prior of `msBayes` is in generations, thus the logic of Hickerson et al. (2013) requires the additional assumption that all 22 Philippine taxa have a generation time of one year. Thus, the argument of Hickerson et al. (2013) that divergence time estimates of Oaks et al. (2013) "should set an upper bound on their prior for τ" seems questionable, especially given our findings on the behavior of empirically informed priors presented above.

Finally, even *if* we make all of these assumptions and assume the gene divergence times are estimated without error, the probabilistic argument only applies to one of the three different priors used by Oaks et al. (2013). The narrowest prior on divergence times used by Oaks et al. (2013) closely mirrors the range of estimates of gene-divergence times, and applying Hickerson et al.'s (2013) probability equations demonstrates that the prior is densely populated with samples with large numbers of divergence parameters that are consistent with the gene divergence estimates. Thus, according to the argument of Hickerson et al. (2013), if insufficient prior sampling is to blame for the bias, it should be much reduced under the narrow prior on τ. However, the magnitude of the bias is very similar across all three priors explored by Oaks et al. (2013). Hickerson et al. (2013) point out a case where the narrow prior performs slightly better (panel L of Figures S32, S37, and S38 of Oaks et al. (2013)). However, it is important to note that these results suffered from a bug in `msBayes`, and there are many cases after Oaks et al. (2013) corrected the bug where the narrow prior performs slightly worse (see panels D–J of Figures 3 and S12).

To disentangle whether model likelihoods or insufficient prior sampling is to blame for the biases revealed by Oaks et al. (2013), we must look at the different predictions made by

these two phenomena. One example, as discussed by Oaks et al. (2013), is that insufficient prior sampling should create higher variance in posterior estimates, and thus it should cause analyses to be sensitive to the number of samples drawn from the prior. Oaks et al. (2013) do not see such sensitivity when they compare prior sample sizes of 2×10^6, 5×10^6, and 10^7.

To explore this prediction further, we repeat the analysis of the Philippines dataset under the intermediate prior used by Oaks et al. (2013) ($\tau \sim U(0, 10)$, $\theta_D \sim (0.0005, 0.04)$, $\theta_A \sim (0.0005, 0.02)$), using a very large prior sample size of 10^8. When we look at the trace of the estimates of the dispersion index of divergence times (Ω) as the prior samples accumulate (Figure 2.12) we see no trend in either the unadjusted or GLM-regression-adjusted estimates. This suggests that insufficient prior sampling did not play a large role in the bias found by Oaks et al. (2013).

Our results presented above that demonstrate the bias of the model-averaging approach of Hickerson et al. (2013) both toward models with narrower τ priors (Table 2.1 and Figs. 2.3, 2.4, 2.9 and 2.10) and models with fewer τ parameters (Figs. 2.7 & 2.8) strongly suggest the primary cause is the uniform priors reducing the marginal likelihoods of large models. In all of our model-averaging analyses, all prior models have the same number of samples. While analyses that sample each model proportional to their relative parameter space could be explored, it seems much more likely that the broad uniform priors are simply inhibiting the marginal likelihoods of spacious models.

As discussed by Oaks et al. (2013), a straightforward prediction of reduced likelihoods of complex models is that they should disappear as the model generating the data converges to the prior. Oaks et al. (2013) test this prediction by performing 100,000 simulations to assess the model-choice behavior of `msBayes` when the prior model is correct. They find that the bias actually switches direction (at least for the regression-adjusted estimates) and the method tends to underestimate the probability of the model with one divergence event (see Figure 4 of Oaks et al. (2013)). The same prediction is not as straightforward for the insufficient-sampling hypothesis. Even when the prior is correct, due to the discrete

uniform prior on Ψ implemented in `msBayes`, models with larger numbers of divergence events (and thus greater parameter space) will still be less densely sampled than those with fewer divergence events (Oaks et al., 2013). Thus, the results of the simulations of Oaks et al. (2013) are more consistent with marginal model likelihoods causing the bias toward models with fewer divergence time parameters.

Taken together, the results presented here and in Oaks et al. (2013) support reduced marginal likelihoods of large models as a primary mechanism by which broad uniform priors cause biases in `msBayes`. Nonetheless, posterior sampling error will always be present in any numerical Bayesian approximation method. Thus, insufficient sampling of the prior will contribute to the error of all approximate Bayesian estimates. However, there is no strong evidence that it is playing a large role in the biases revealed herein and by Oaks et al. (2013).

2.7 Validation analyses

Following Oaks et al. (2013), we characterize the model-choice behavior of the model-averaging approach of Hickerson et al. (2013) under the ideal conditions where the prior is correct (i.e., the data are generated from parameters drawn from the same prior distributions used in the analysis). We used the same prior models as above (M_1–M_5; Table 2.1), and generated 50,000 pseudo-replicate datasets under this prior (10,000 from each model). We used a simulated data structure of eight population pairs, with a single 1000 base-pair locus sampled from 10 individuals from each population. We then analyzed each of these replicate datasets using the same prior, retaining 1000 posterior samples. Our results are very similar to Oaks et al. (2013), but we note that they are not directly comparable as our simulations contained eight population pairs rather than 10 (Figure 2.11). We find that the approach of Hickerson et al. (2013) estimates the posterior probability of divergence models reasonably well when all assumptions of the method are met (i.e., the prior is correct) and the unadjusted posterior estimates are used. Similar to Oaks et al. (2013), we find that the

regression-adjusted estimates of the model probabilities are biased.

2.8 Differing utilities of Ψ and Ω in `msBayes`

The main parameter of the `msBayes` model is the vector of divergence times for each of the taxon pairs, $\boldsymbol{\tau} = \{\tau_1, \ldots, \tau_Y\}$ (Oaks et al., 2013). Hickerson et al. (2013) argue that the dispersion index of this vector, Ω, is a better model-choice estimator than the number of divergence time parameters within the vector, Ψ. They present a plot of Ψ against Ω (Fig. S1 of Hickerson et al. (2013)), which is simply a plot of sample size versus variance. This plot shows, not surprisingly, that Ω has very little information about the number of divergences among taxa. Nonetheless, Hickerson et al. (2013) conclude Ω more informative and biogeographically relevant than Ψ. Certainly the maximum information is contained within the divergence time vector that Ω is summarizing. Clearly, the number of divergence time parameters within the vector and their values is more informative than its variance (i.e., the dispersion index is not a sufficient statistic for $\boldsymbol{\tau}$).

Hickerson et al. (2013) also argue that "`msBayes` can estimate Ω much better than Ψ." However, `msBayes` is a model-choice estimator, and hence the goal is to estimate the posterior probability of divergence models. Oaks et al. (2013) demonstrate that even when all assumptions of the model are met, Ω_is a poor model-choice estimator (see plots B, D & F of Figure 4), whereas Ψ performs better. Furthermore, Ω is limited to estimating the probability of only a single divergence model (the one divergence model), and thus its utility for model choice is extremely limited.

The model-choice utility of Ω is limited to the probability that this continuous statistic, which can range from zero to infinity, is at its lower limit of zero. In theory, this point density will always be zero, thus an arbitrary threshold (0.01 is used throughout the `msBayes` literature) must be chosen to make the probability estimable. However, it is still not surprising to see that it is numerically difficult to obtain reliable estimates of the probability that the

continuous Ω statistic is "near" its limit of zero. It is much easier, less subjective, and more interpretable to estimate the probability of the discrete parameter of the model, Ψ, is at its lower limit of one. Thus, it is not surprising that Oaks et al. (2013) find that Ψ is a better estimator of model probability than Ω.

2.9 Some general thoughts on the model of `msBayes`

Our results demonstrating the poor behavior of the model-averaging approach proposed by Hickerson et al. (2013) are not too surprising. While this approach is trivial to implement, it is not a trivial change to the `msBayes` model. To better understand the reason for this, it might help to step back and get a sense of the scale of the `msBayes` model. To do this, we will use the dataset of 22 vertebrate taxon pairs of Oaks et al. (2013) as an example.

Following the model description and notation of Oaks et al. (2013), let us tally up all of the free parameters in the `msBayes` model. Under the simplest model in `msBayes` (i.e., assuming no migration and no intra-locus recombination), the number of parameters for each taxon pair include: The population sizes of the ancestral and descendant populations (θ_A, θ_{D1}, θ_{D2}), the magnitude of population contraction in each of the descendant populations (ζ_{D1} and ζ_{D2}) and the timing of these contractions (τ_B), and the $N-1$ node heights (coalescent times) of the gene tree that gave rise to the N gene copies sampled from both populations of the pair. Because we only have a single locus for each taxon, there are no locus-specific θ-scaling parameters (υ) or α shape parameter for the gamma prior distribution on υ. Lastly, there are between one and 22 divergence time parameters τ in the vector $\boldsymbol{\tau}$. Overall, for our Philippines data under the simplest model in `msBayes`, there are 581–602 free parameters (depending on the cardinality of $\boldsymbol{\tau}$). Furthermore, under this rich model, the method is sampling over 1002 divergence models (i.e., the number of integer partitions of $Y = 22$; Oaks et al., 2013).

This is a very difficult inference problem, and the method only uses four summary statis-

tics calculated from the sequence alignment of each taxon pair: π (Tajima, 1983), θ_W (Watterson, 1975), π_{net} (Takahata and Nei, 1985), and $SD(\pi - \theta_W)$ (Tajima, 1989). That gives us a total of 88 summary statistics (four from each of the 22 taxon pairs), which contain minimal information about many of the $\approx$ 600 parameters in the model. More summary statistics can be used in `msBayes`, but most are highly correlated with these four statistics (which are even highly correlated among themselves), and thus contribute little additional information about the data.

The large number of parameters and divergence models relative to summary statistics is undoubtedly a key reason the method is so sensitive to the prior distributions. It also likely contributes to the method's lack of robustness. Robustness is an extremely important characteristic of a method to gauge its applicability to real-world data, because we know the model and priors will be wrong to some degree.

The approach of Hickerson et al. (2013) adds an additional dimension of model choice to the model. They expand the model to sample over eight prior models. This extends the original model to having 582-603 free parameters and, more importantly, sampling over 8016 unique model states across both divergence models and prior models. This is a non-trivial extension of the model and, given the method's very large number of parameters and divergence models relative to the information used from the data, likely plays a major role in the poor behavior of this approach.

In theory, the approach of Hickerson et al. (2013) is very appealing. It sums over multiple candidate prior models to produce a posterior estimate marginalized over the uncertainty in prior choice. In general, Bayesian model-averaging is a powerful approach that leverages a great strength of Bayesian statistical procedures, namely the ability to obtain marginalized estimates that incorporate uncertainty in nuisance parameters. However, given the basic `msBayes` model is already struggling to estimate a huge number of parameters and model probabilities with scant information from the data, it is not surprising that adding another dimension of model choice to the method causes problems.

The recommendations of Oaks et al. (2013) for mitigating the bias and lack of robustness of `msBayes` are actually similar to those of Hickerson et al. (2013), but avoid the need for imposing additional model choice. Oaks et al. (2013) suggest that uniform priors are inappropriate for many parameters of the `msBayes` model, and recommend the use of probability distributions from the exponential family. If we look at the prior distribution on the divergence time parameter τ imposed by the model-averaging approach of Hickerson et al. (2013) we see it is a mixture of overlapping uniforms with lower limits of zero (Figure 2.13). This looks very much like an exponential distribution, except that each sample of divergence times are restricted to the hard bounds of one of the eight prior models. Thus, it seems more appropriate to place a gamma prior (the exponential being a special case of the gamma) on divergence times. This would provide investigators flexibility to represent their prior uncertainty in model parameters while avoiding broad uniform distributions and the need of costly model-averaging. This seems like a natural solution to the problem of skewed marginal likelihoods and/or sampling error, which are both likely exacerbated by the fact that uniform distributions are the only choice of prior for many of the parameters of the `msBayes` model (τ, θ_A, θ_{D1}, θ_{D2}, τ_B, ζ_{D1}, ζ_{D2}, α, m, r; Oaks et al., 2013).

Given that ABC methods only need to draw random values from prior distributions (hence there are no difficult prior and proposal ratios to calculate, etc.) it is easy to use the most appropriate distributions on parameters. Also, this approach reduces the temptation of empirically guided priors. The investigator can place the majority of the prior density in regions of parameter space they believe, *a priori*, are most plausible, but still capture uncertainty in the tails of distributions with low density. As discussed in Oaks et al. (2013) the use of uniform priors necessitate broad priors to avoid excluding the truth *a priori*. Thus, to avoid the behavior of the method under these uniform priors, using the data is a very tempting, albeit risky, option (Hickerson et al., 2013). More appropriate prior distributions would alleviate this issue.

2.10 Other clarifications

2.10.1 Graphical checks of priors

Hickerson et al. (2013) advocate the use of what they call graphical checks of prior models. This entails generating a small number of random samples from the prior (1000) and plotting the resulting summary statistics in comparison to the observed statistics to see if they coincide (see Figure 1 of Hickerson et al. (2013)). As we show above, this is a risky strategy, because the resulting plots of this approach have little correlation with the appropriateness of priors. Given the richness of the `msBayes` model ($\approx$ 600 parameters for the Philippine dataset analyzed by Hickerson et al. (2013)), we do not expect that 1000 *random* draws from the vast prior parameter space will yield data and summary statistics consistent with the observed data. In fact, when the random draws are tightly clustered around the observed statistics, this can be an indication that the prior is over-fit, as we show above (Table 2.1 and Figure 2.2). Thus, interpreting such plots should be avoided, and the use of posterior predictive analyses would be much more informative about the overall fit of models.

2.10.2 The validity of `msBayes` estimates

Hickerson et al. (2013) claim that Oaks et al. (2013) "assume that all previous `msBayes` results are invalid." With the exception of Hickerson et al. (2013), we do not claim any previous results of `msBayes` are invalid. Rather, we conservatively recommend that the common inference of temporally clustered divergences (Barber and Klicka, 2010; Bell et al., 2012; Carnaval et al., 2009; Chan et al., 2011; Daza et al., 2010; Hickerson et al., 2006; Huang et al., 2011; Lawson, 2010; Leaché et al., 2007; Plouviez et al., 2009; Stone et al., 2012; Voje et al., 2009), when not accompanied with the necessary simulation-based analyses to guide the interpretation of such results, should be treated with caution, because the method has been shown to spuriously infer clustered divergences over a range of prior conditions.

2.10.3 Saturation of summary statistics

Hickerson et al. (2013) claim that our priors "cause much of the explored parameter space to be beyond the threshold of saturation in most mtDNA genes." To explore this possibility, we simulated datasets under prior settings that match two of the three priors used by Oaks et al. (2013): $\theta_D \sim U(0.0005, 0.04)$ and $\theta_A \sim U(0.0005, 0.02)$. Under this prior, we draw divergence time parameters from a uniform distribution of $U(0, 20)$, simulate datasets, and plot the τ values against the summary statistics calculated from the resulting datasets (Figure 2.14). Clearly, the priors used by Oaks et al. (2013) with upper limits on τ of five and 10 suffered from little to no effect from saturation. Even at divergence times of 20 coalescent units, there is still signal in the summary statistics used by `msBayes` (Figure 2.14). Thus, the assertion of Hickerson et al. (2013) does not apply to at least two of the priors used by Oaks et al. (2013) and, as a result, does not explain the bias they found.

2.11 Conclusions

We demonstrate how the approximate Bayesian model choice method implemented in `msBayes` can be strongly biased away from models with greater parameter space. As suggested by Oaks et al. (2013), this is likely caused by the inappropriate use of uniform priors on most of the model's parameters. These priors necessitate the use of broad priors that place high prior density in unlikely regions of parameter space, less the risk of excluding the truth *a priori*. This likely reduces the marginal likelihoods of models with greater parameter space, either due to more divergence time parameters or broader prior distributions on those parameters. We show that the empirical Bayesian model-averaging approach of Hickerson et al. (2013) does not mitigate this bias, but rather causes it to manifest by sampling predominantly from models that may exclude the true values of the parameters. Exploring alternative prior probability distributions on most of the model's parameters, in addition to different priors over divergence models, should help mitigate the method's biases.

The work presented herein follows the principles of Open Notebook Science. All aspects of the work were recorded in real-time via version-control software and are publicly available at https://github.com/joaks1/msbayes-experiments. All information necessary to reproduce our results are provided there.

2.12 Tables

Table 2.1. Results of the model-averaging approach of Hickerson et al. (2013) applied to the Philippines dataset of Oaks et al. (2013) using three sets of prior models. All models used priors on population size of $\theta_D \sim U(0.0001, 0.1)$ and $\theta_A \sim U(0.0001, 0.05)$, and differ only in their prior on divergence time (τ) parameters. Each set of five models differ only in the divergence time prior used for the model with the narrowest prior: M_1 ($\tau \sim U(0, 0.1)$), M_{1A} ($\tau \sim U(0, 0.01)$), or M_{1B} ($\tau \sim U(0, 0.001)$). The approximate posterior probability of each model ($p(M_i \mid B_\epsilon(\mathbf{S}^*))$) is given for each of the three analyses. The posterior estimates are based on 10,000 samples retained from 1×10^6 prior samples from each model.

		$p(M_i \mid B_\epsilon(\mathbf{S}^*))$		
Model	τ prior	$M_* = M_1$	$M_* = M_{1A}$	$M_* = M_{1B}$
M_*	–	0.899	0.821	0.673
M_2	$U(0, 1)$	0.079	0.136	0.251
M_3	$U(0, 5)$	0.013	0.026	0.044
M_4	$U(0, 10)$	0.006	0.012	0.022
M_5	$U(0, 20)$	0.003	0.005	0.010

2.13 Figures

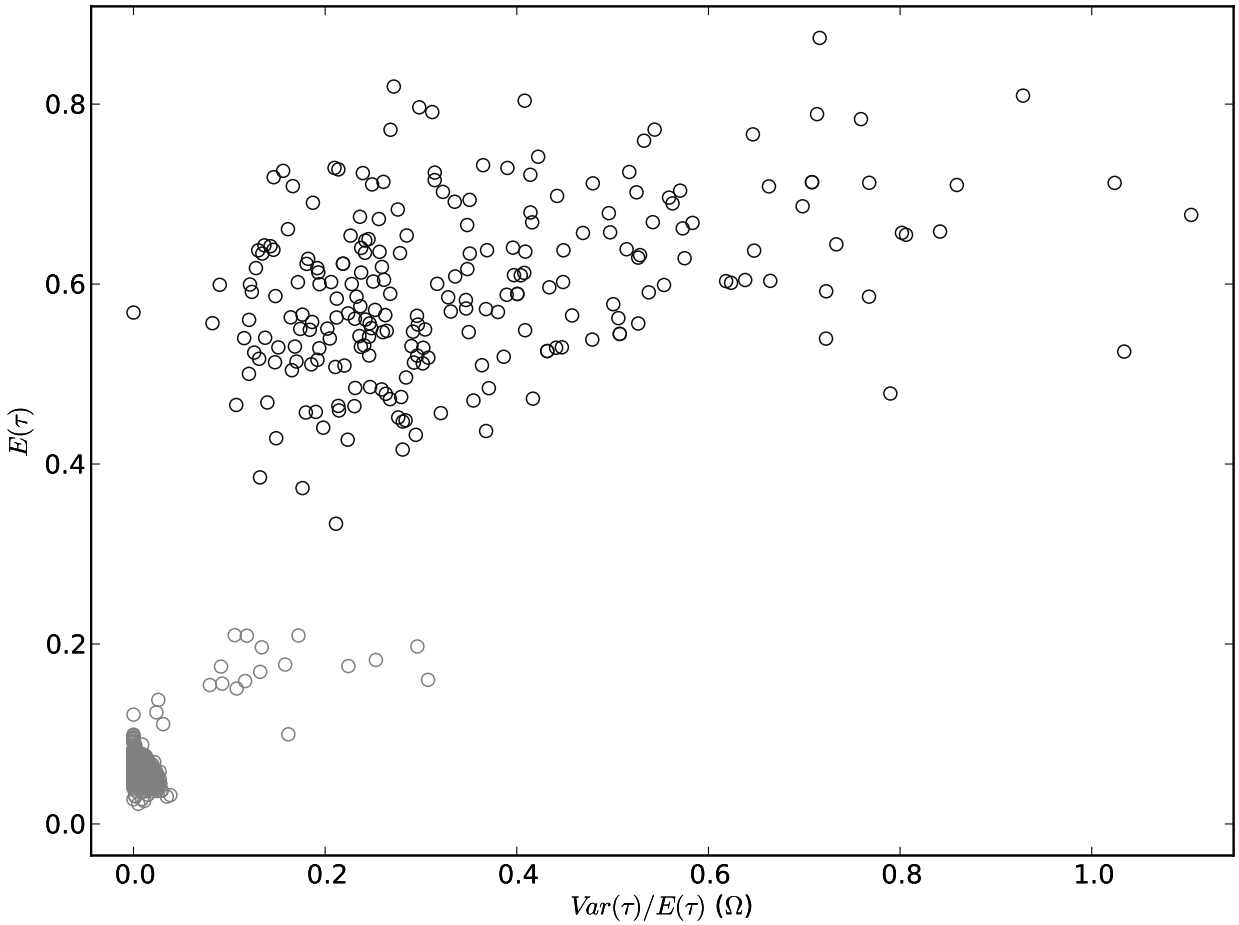

Figure 2.1. The joint posterior of the mean ($E(\tau)$) and dispersion index ($\Omega = Var(\tau)/E(\tau)$) of divergence times for 22 vertebrate taxon pairs as estimated by Hickerson et al. (2013) (see Figure 2B of Hickerson et al. (2013)). The posterior samples are color-coded to indicate the erroneous mixture of timescales in the analysis of Hickerson et al. (2013); grey = $0.05/\mu$ generations and black = $0.02/\mu$ generations.

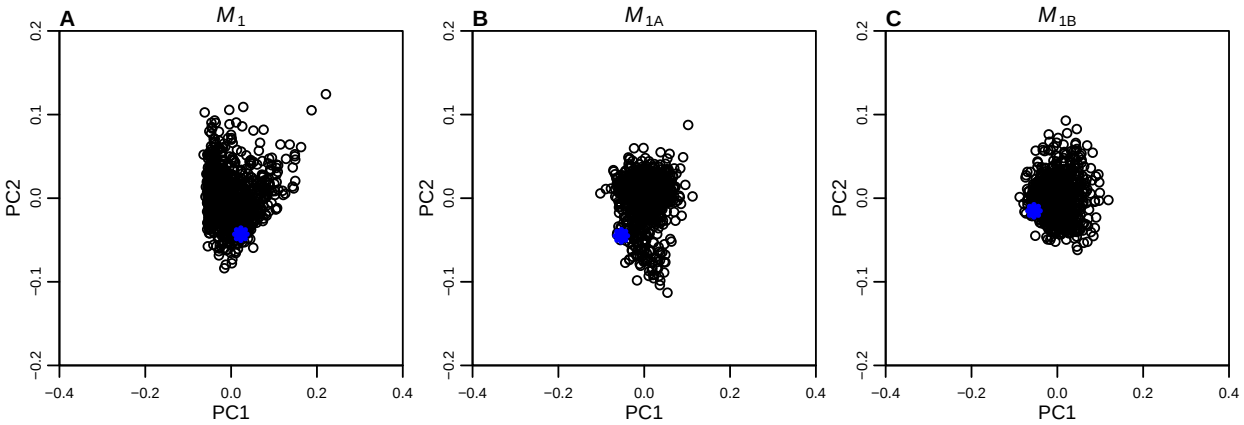

Figure 2.2. The graphical checks recommended by Hickerson et al. (2013) for three prior models: (A) M_1 ($\tau \sim U(0, 0.1)$), (B) M_{1A} ($\tau \sim U(0, 0.01)$), and (C) M_{1B} ($\tau \sim U(0, 0.001)$). The plots project the summary statistics from 1000 random samples from each model onto the first two orthogonal axes of a principle component analysis, with the blue dot representing the observed summary statistics from the 22 population pairs of Philippine vertebrates.

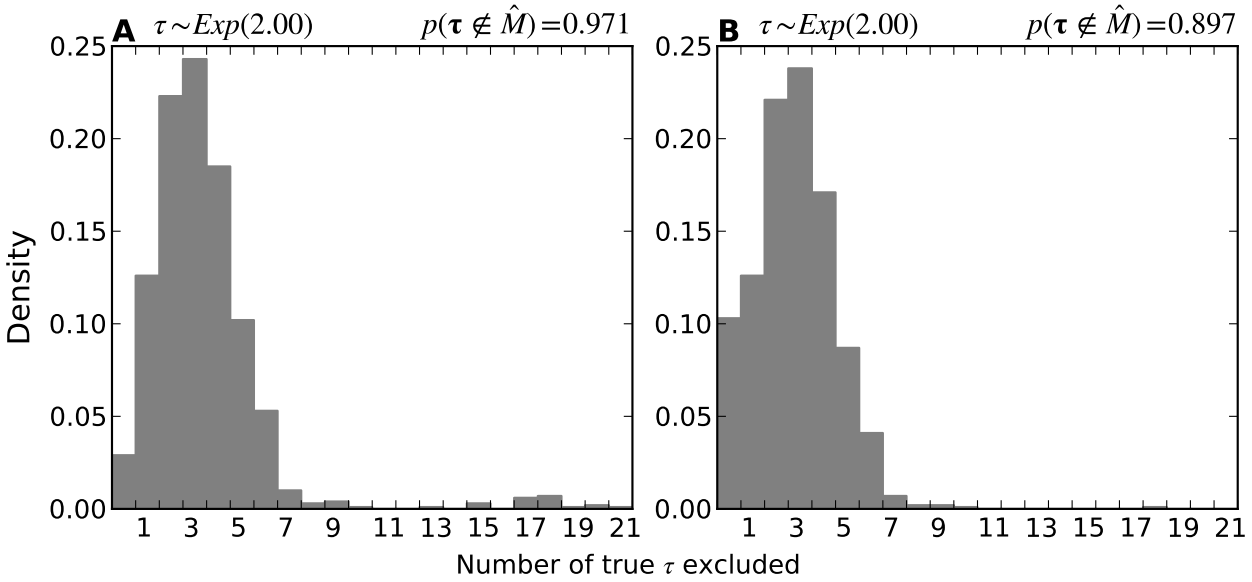

Figure 2.3. The propensity of the model-averaging approach of Hickerson et al. (2013) to exclude the truth. The plots illustrate the number of true τ parameters excluded from analyses of simulated datasets where τ for 22 pairs of populations is drawn from an exponential distribution, $\tau \sim Exp(2)$. The plots represent (A) unadjusted and (B) GLM-adjusted estimates from 1000 simulation replicates analyzed using 5×10^6 samples from the prior. The proportion of simulation replicates in which at least one true parameter value is excluded from the model preferred by a Bayes factor ($p(\tau \notin \hat{M})$) is also given.

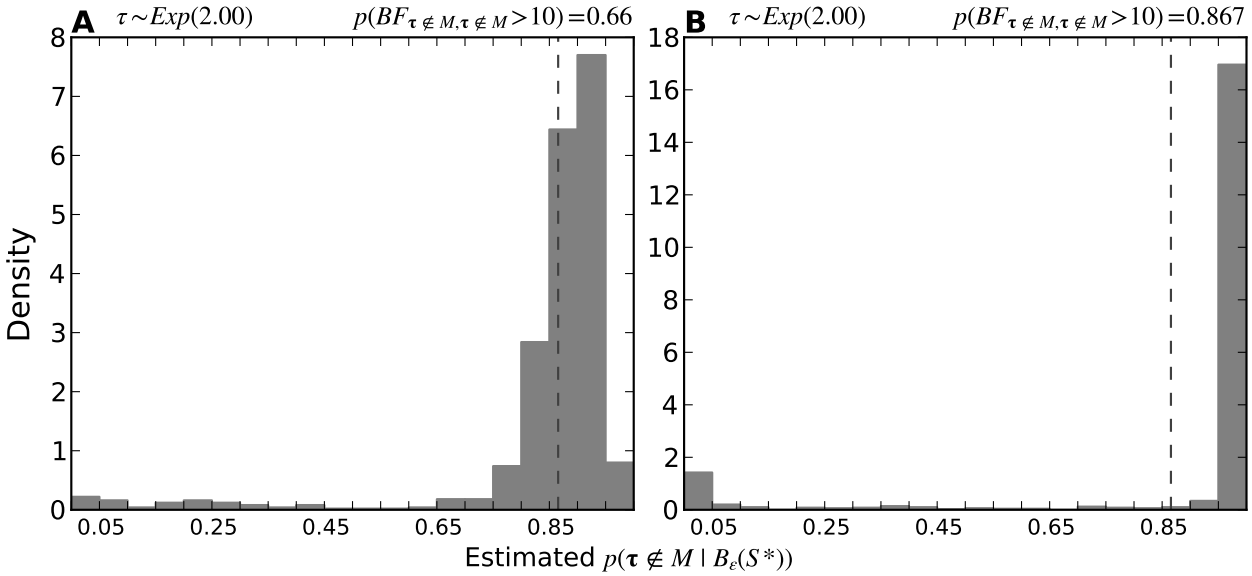

Figure 2.4. The propensity of the model-averaging approach of Hickerson et al. (2013) to exclude the truth. The plots illustrate the estimated probability of excluding at least one true τ value from analyses of simulated datasets where τ for 22 pairs of populations is drawn from an exponential distribution, $\tau \sim Exp(2)$. The plots represent (A) unadjusted and (B) GLM-adjusted estimates from 1000 simulation replicates analyzed using 5×10^6 samples from the prior. The proportion of simulation replicates in which there is strong support for at least one true parameter value being excluded from the model ($p(BF_{\tau \notin M, \tau \in M} > 10)$) is also given.

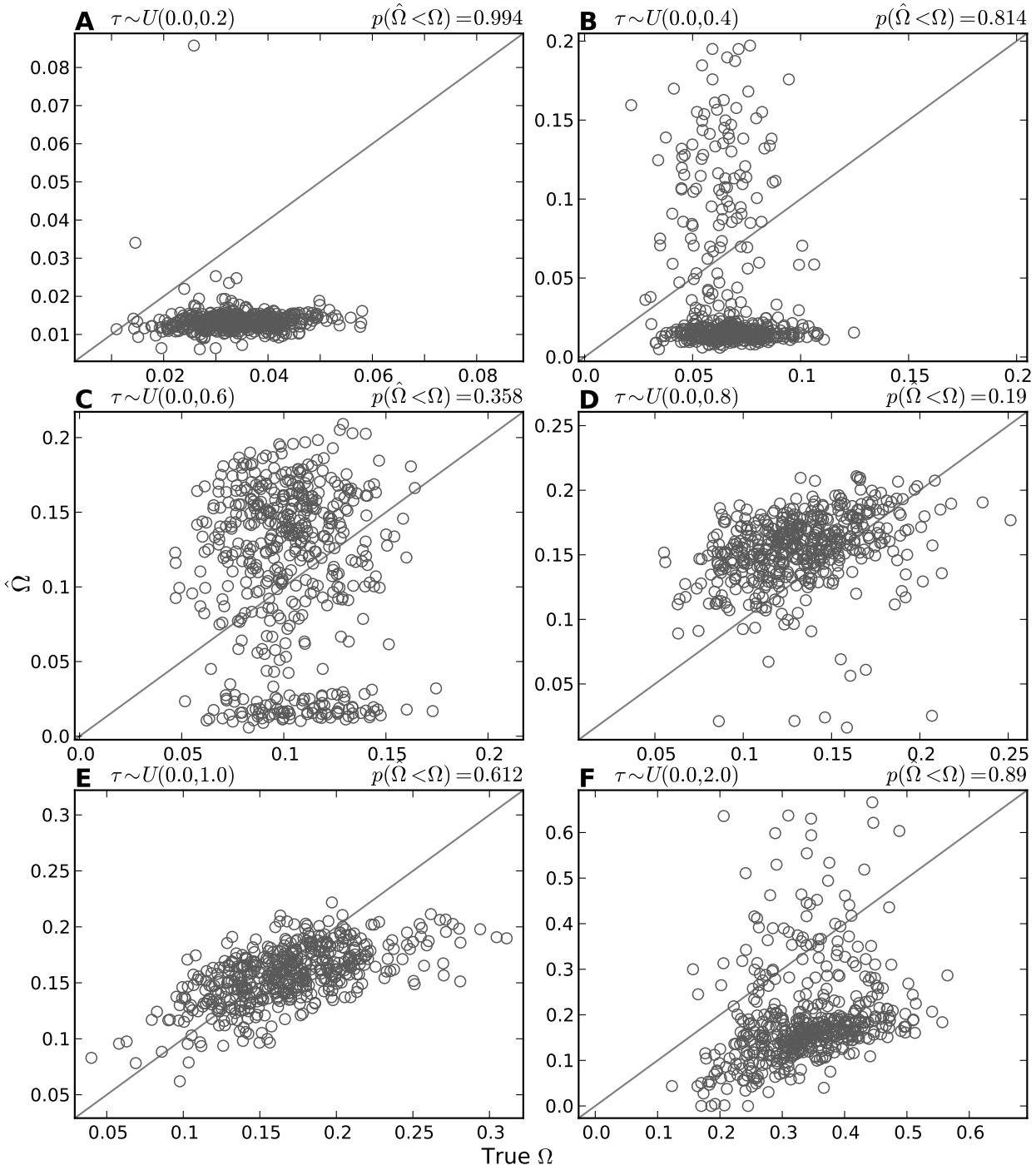

Figure 2.5. The accuracy of the model-averaging approach of Hickerson et al. (2013) to estimate the dispersion index of divergence times (Ω) from analyses of simulated datasets where τ for 22 pairs of populations is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value of, $p(\hat{\Omega} < \Omega)$, is given for each τ_{max}. Each plot represents unadjusted median estimates from 500 simulation replicates analyzed using 5×10^6 samples from the prior.

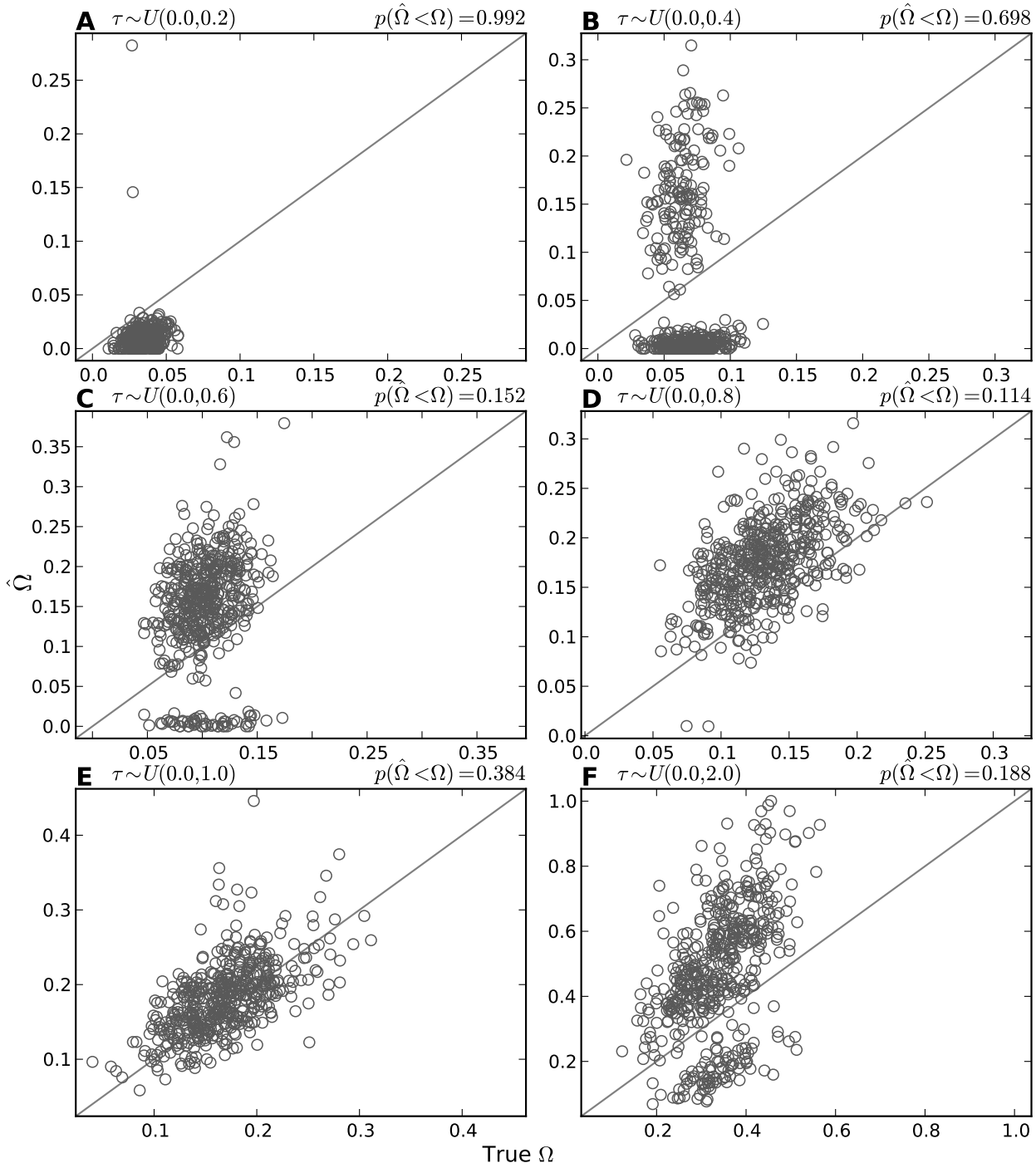

Figure 2.6. The accuracy of the model-averaging approach of Hickerson et al. (2013) to estimate the dispersion index of divergence times (Ω) from analyses of simulated datasets where τ for 22 pairs of populations is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value of, $p(\hat{\Omega} < \Omega)$, is given for each τ_{max}. Each plot represents GLM-regression-adjusted mode estimates from 500 simulation replicates analyzed using 5×10^6 samples from the prior.

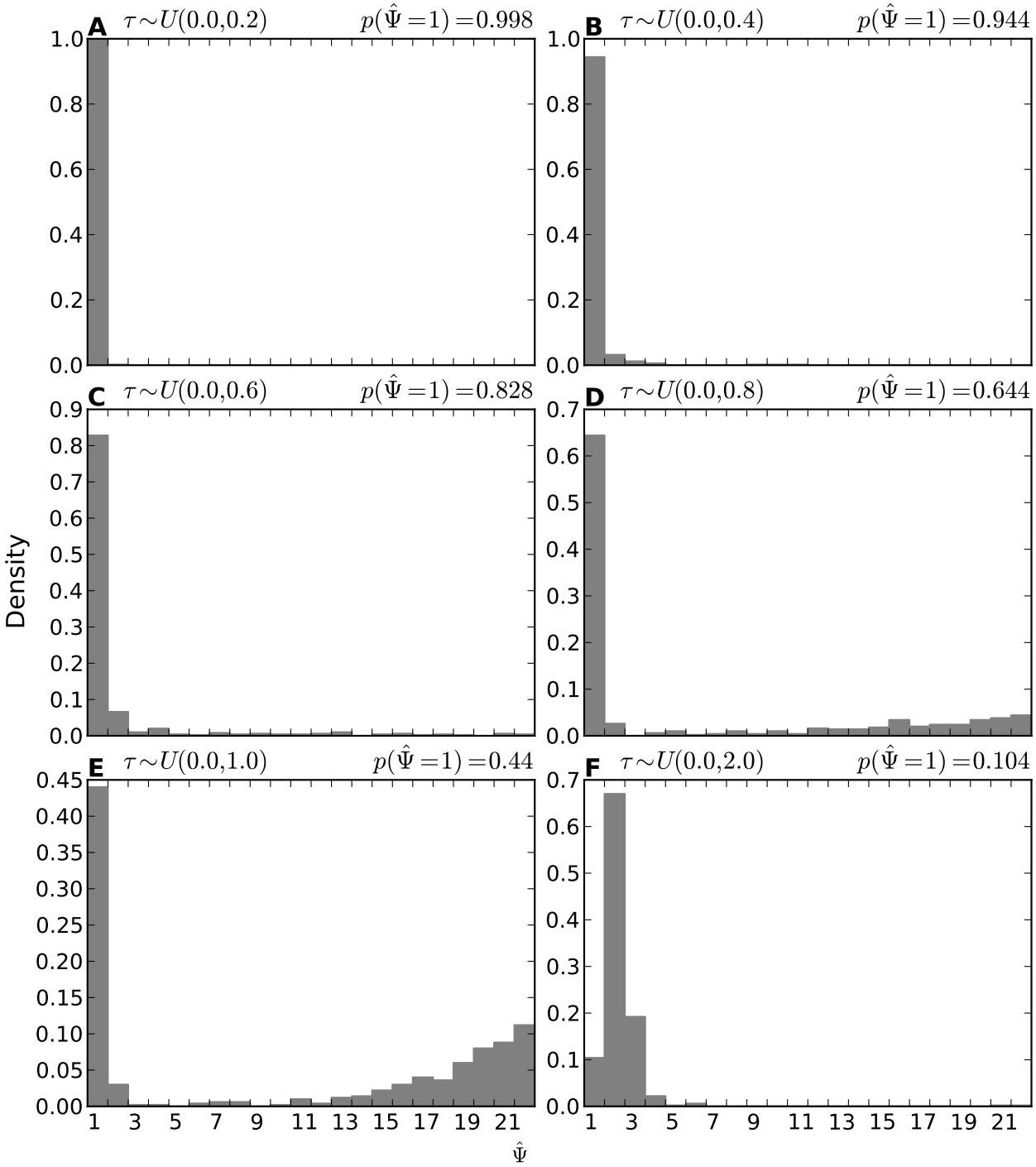

Figure 2.7. The bias of the model-averaging approach of Hickerson et al. (2013) to infer clustered divergences in our simulation-based power analyses. The plots illustrate the estimated number of divergence events ($\hat{\Psi}$) from analyses of simulated datasets where τ for 22 pairs of populations is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of the method inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. Each plot represents 500 simulation replicates analyzed using 5×10^6 samples from the prior.

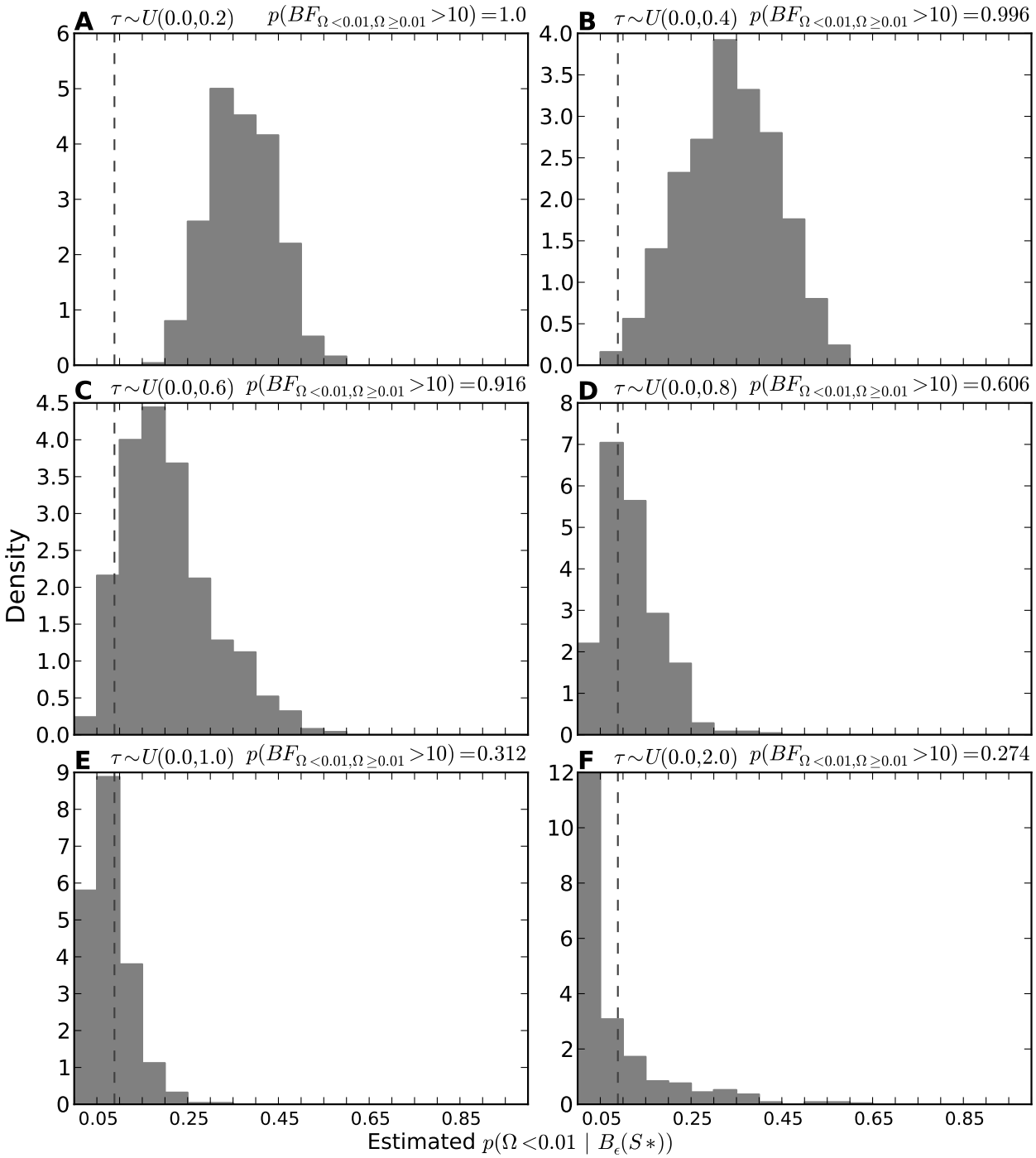

Figure 2.8. The bias of the model-averaging approach of Hickerson et al. (2013) to support one divergence event in our simulation-based power analyses. The plots illustrate histograms of the estimated posterior probability that the dispersion index of divergence times is less than 0.01 ($p(\Omega < 0.01|B_\epsilon(\mathbf{S}^*))$) from analyses of simulated datasets where τ for 22 pairs of populations is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of simulation replicates that strongly support one divergence event, $p(BF_{\Omega<0.01,\Omega\geq0.01} > 10)$, is given for each τ_{max}. Each plot represents 500 simulation replicates analyzed using 5×10^6 samples from the prior.

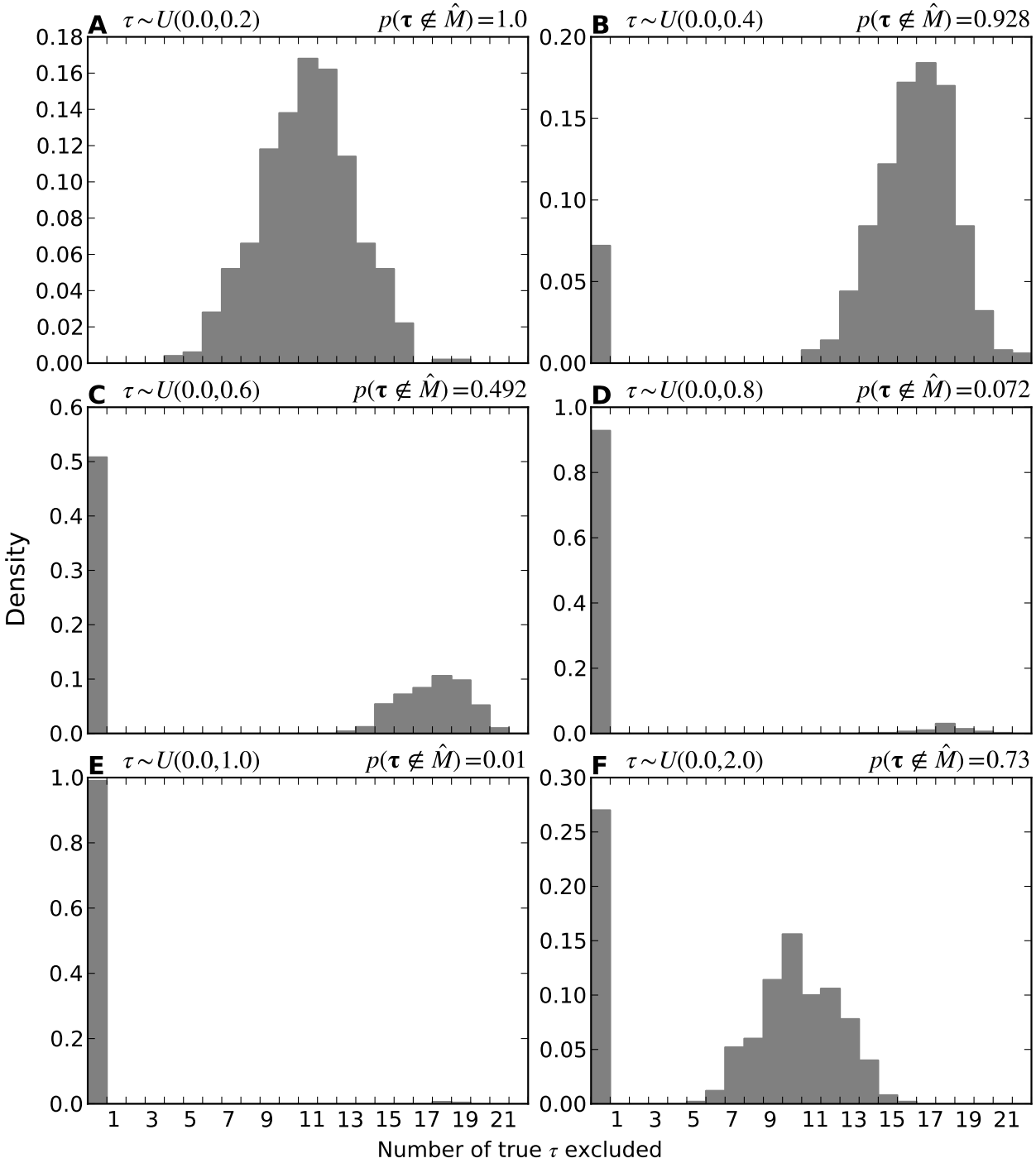

Figure 2.9. The propensity of the model-averaging approach of Hickerson et al. (2013) to exclude the truth in our simulation-based power analyses. The plots illustrate the number of true τ parameters excluded from analyses of simulated datasets where τ for 22 pairs of populations is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of simulation replicates in which at least one true parameter value is excluded from the model preferred by a Bayes factor ($p(\tau \notin \hat{M})$) is given for each τ_{max}. Each plot represents 500 simulation replicates analyzed using 5×10^6 samples from the prior.

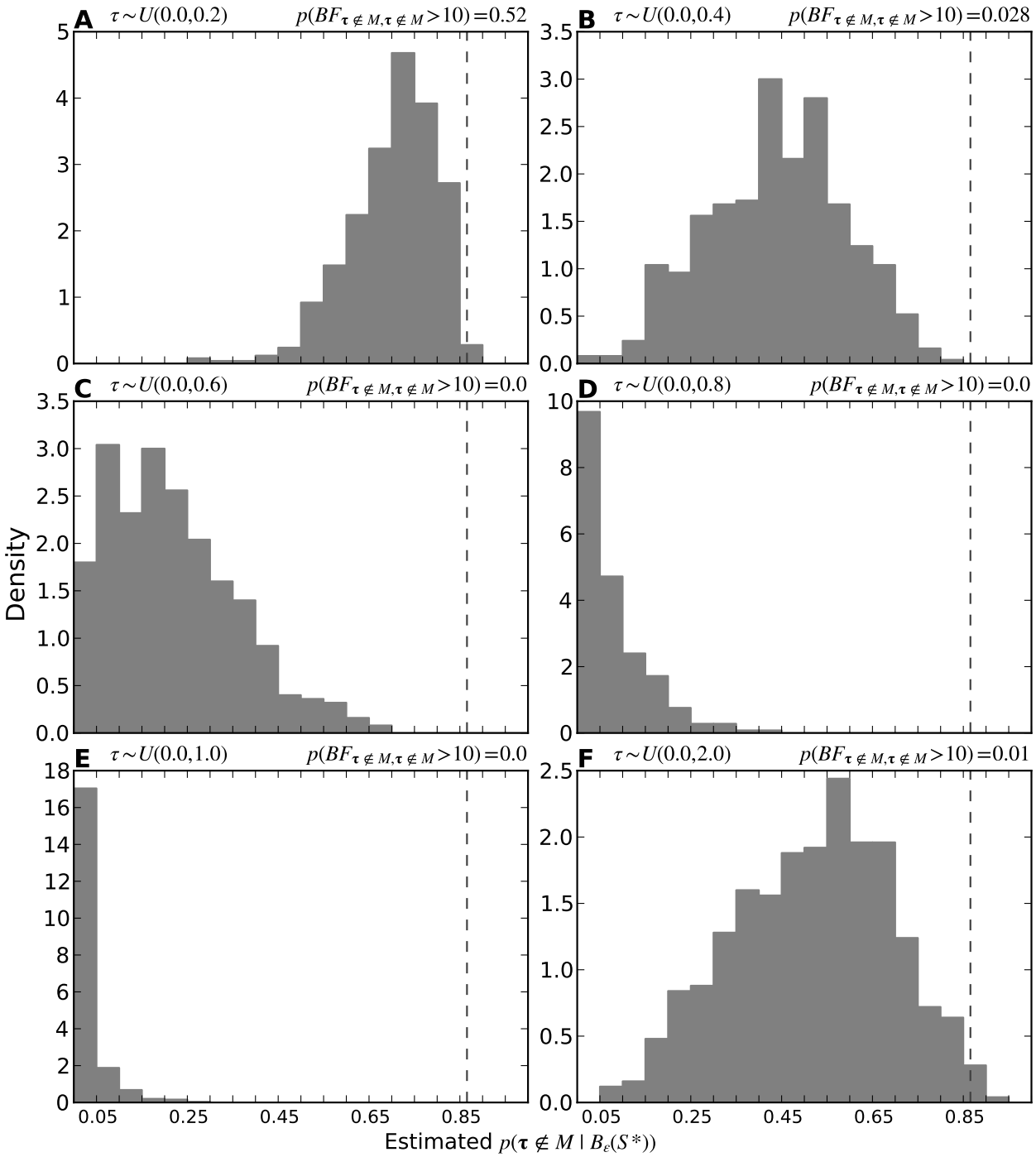

Figure 2.10. The propensity of the model-averaging approach of Hickerson et al. (2013) to exclude the truth in our simulatoin-based power analyses. The plots illustrate the estimated posterior probability of excluding at least one true τ value from analyses of simulated datasets where τ for 22 pairs of populations is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of simulation replicates in which there is strong support for at least one true parameter value being excluded from the model ($p(BF_{\tau \notin M, \tau \in M} > 10)$) is given for each τ_{max}. Each plot represents 500 simulation replicates analyzed using 5×10^6 samples from the prior.

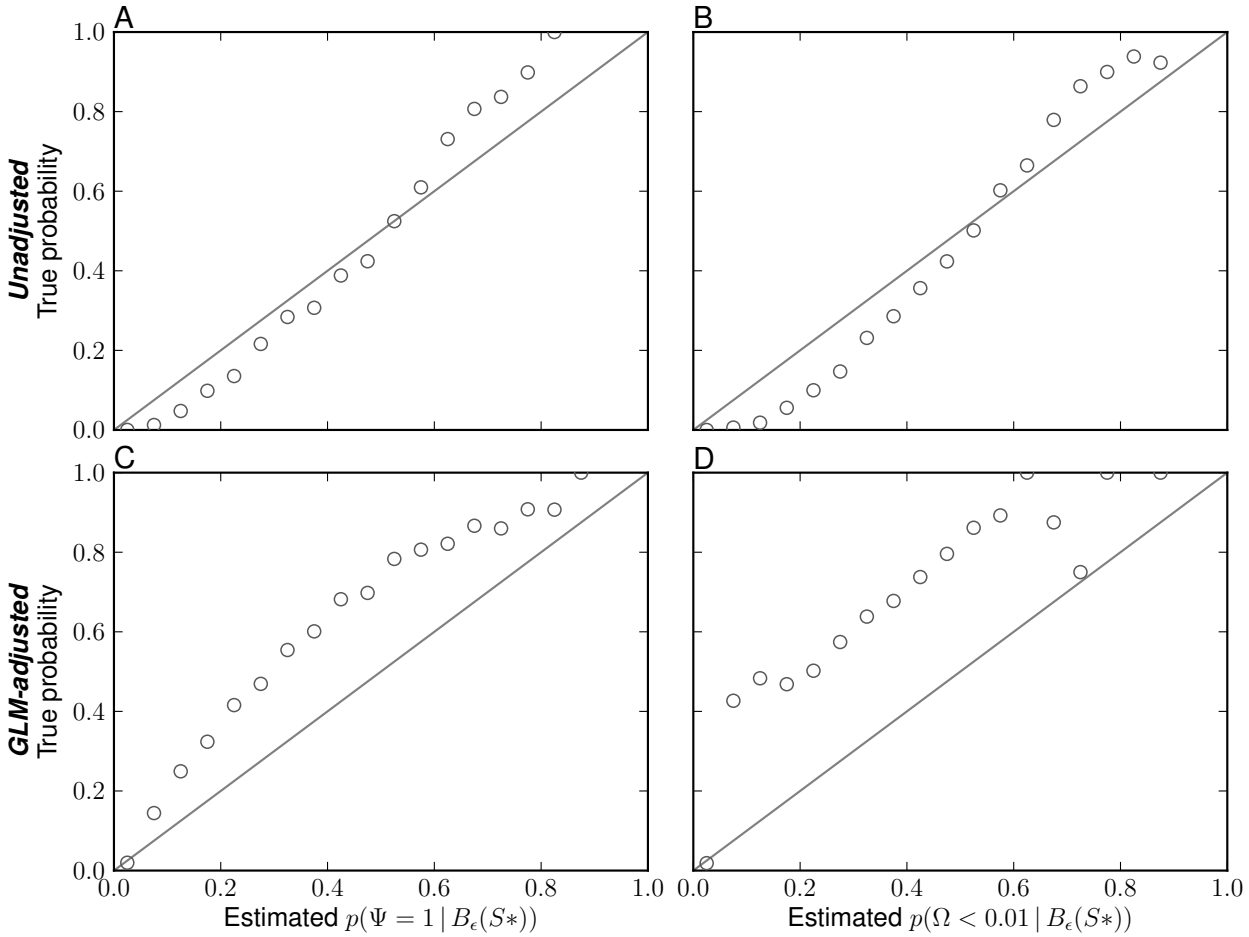

Figure 2.11. An assessment of the approximate Bayesian model-averageing approach of Hickerson et al. (2013) under the ideal conditions when the prior model is correct (i.e., the pseudo-replicate datasets are simulated from parameters drawn from the same prior distributions used in the analysis). The plots show the relationship between the estimated posterior and true probability of (A & C) $\Psi = 1$ and (B & D) $\Omega < 0.01$, based on 50,000 simulations. The results summarize the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimate from each simulation replicate. The prior settings for all replicates included five prior models with $\theta_D \sim U(0.0001, 0.1)$ and $\theta_A \sim U(0.0001, 0.05)$ for all five models, and $M_1 : \tau \sim U(0, 0.1)$, $M_2 : \tau \sim U(0, 1)$, $M_3 : \tau \sim U(0, 5)$, $M_4 : \tau \sim U(0, 10)$, and $M_5 : \tau \sim U(0, 20)$. The number of samples from the prior was 2.5×10^6. The simulated data structure was 8 population pairs, with a single 1000 bp locus sampled from 10 individuals from each population. The 50,000 estimates of the posterior probability of one divergence event were assigned to 20 bins of width 0.05. The estimated posterior probability of each bin is plotted against the proportion of replicates in that bin with a true value consistent with one divergence event (i.e., $\Psi = 1$ or $\Omega < 0.01$).

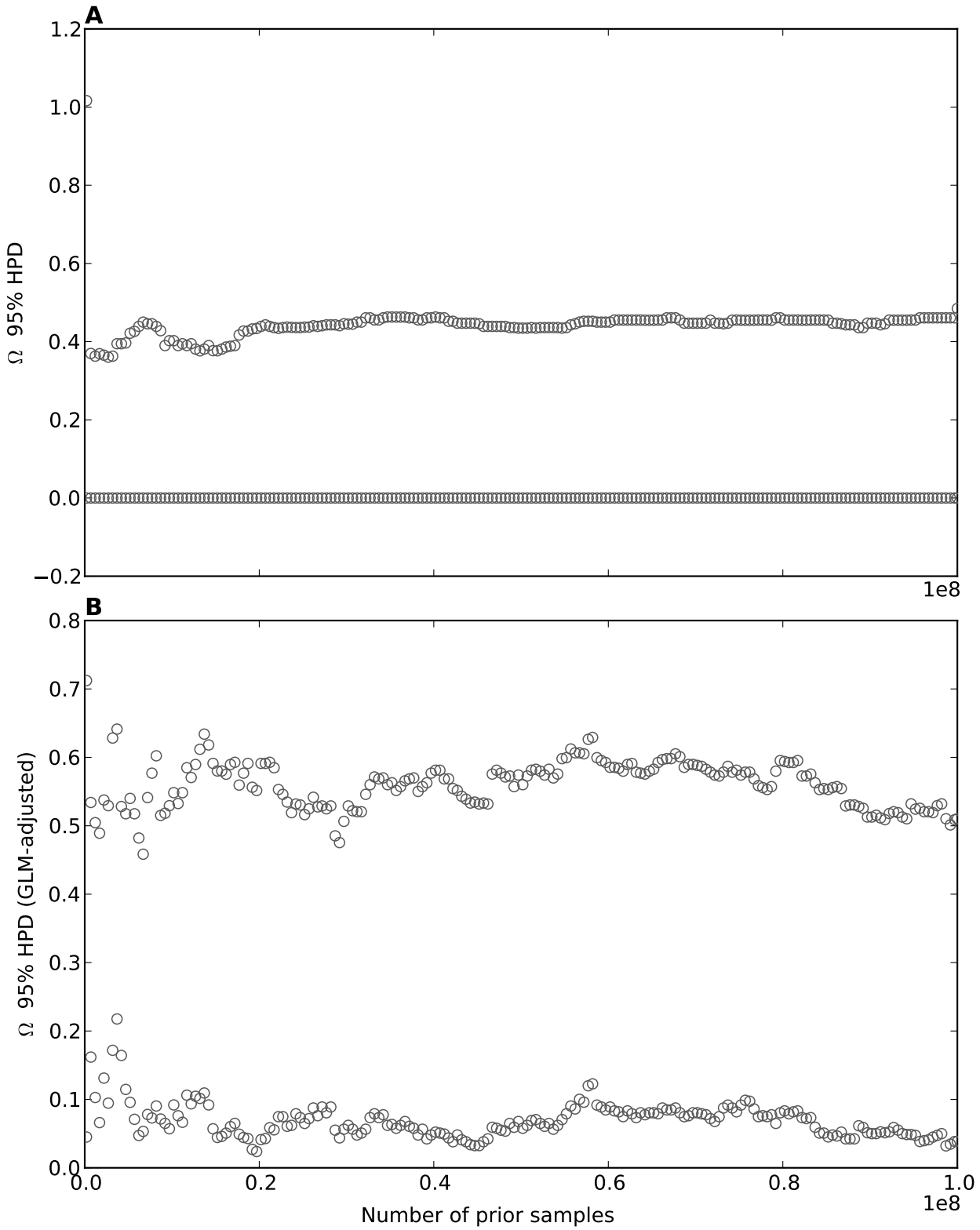

Figure 2.12. Traces of the estimated lower and upper limits of the 95% highest posterior density (HPD) interval of Ω (the dispersion index of divergence times) as 100 million prior samples are accumulated. Each pair of points is based on 1000 posterior samples retained from the prior. Both (A) unadjusted and (b) GLM-regression-adjusted estimates are shown. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

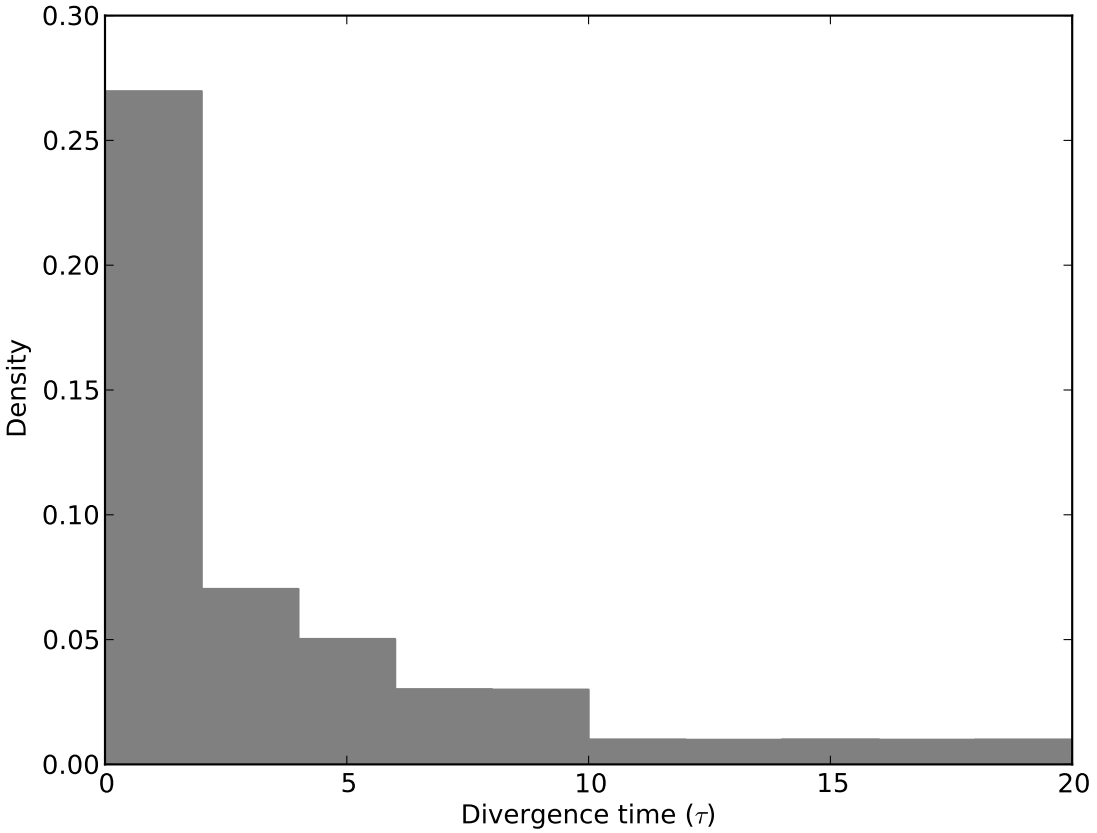

Figure 2.13. The prior distribution on divergence times imposed by the model-averaging prior comprised of five models with different uniform priors on τ: M_1 ($\tau \sim U(0, 0.1)$), M_2 ($\tau \sim U(0, 1)$), M_3 ($\tau \sim U(0, 5)$), M_4 ($\tau \sim U(0, 10)$), M_5 ($\tau \sim U(0, 20)$).

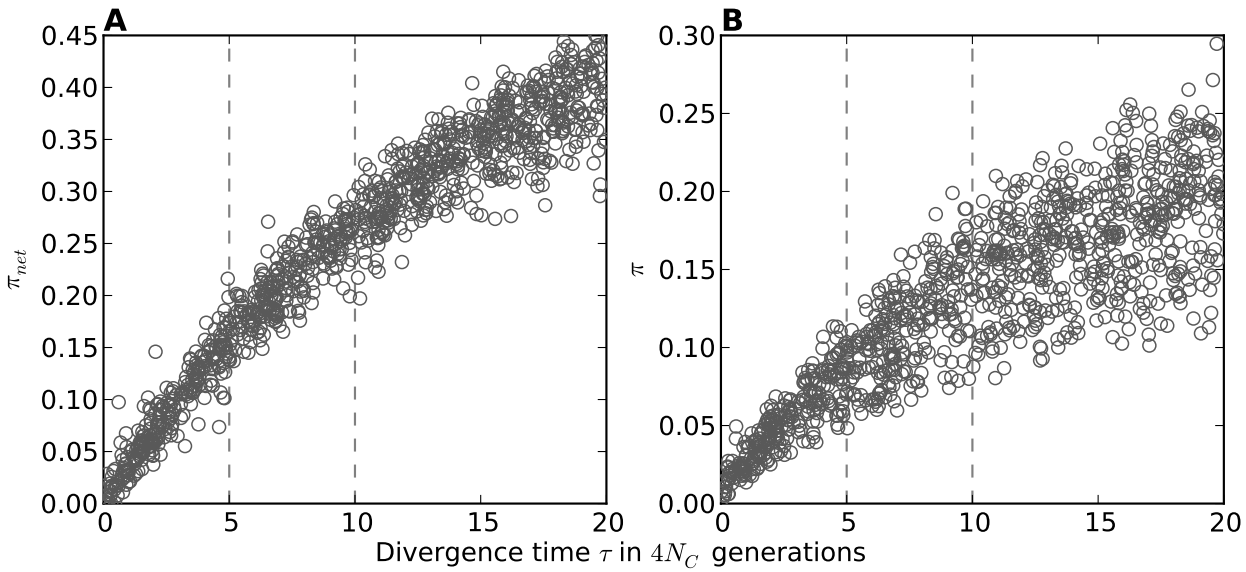

Figure 2.14. The summary statistics π (Tajima, 1983) and π_{net} (Takahata and Nei, 1985) as a function of divergence time between populations. Each plot represents 1100 pairs of parameter draws and summary statistics calculated from the simulated data. Prior settings for the simulations were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

Chapter 3

An Improved Approximate-Bayesian Model-choice Method for Estimating Shared Evolutionary History

3.1 Introduction

Understanding the processes that generate biodiversity and regulate community assembly is a major goal of evolutionary biology. Given the dynamic nature of our planet, a major component of this goal is understanding how geological and climatic processes can affect the evolutionary history of whole groups of co-distributed species and their associated microbiota, including pathogens. Such large-scale processes predict patterns of shared evolutionary history across affected populations of organisms, e.g., temporally clustered divergences across co-distributed species. At recent timescales, such events should leave a signature in the genetic variation within and among the affected lineages. Thus, being able to accurately estimate models of shared evolutionary histories across co-distributed taxa would provide an important tool for better understanding how regional and global biogeographic processes affect biodiversity.

This inference problem is challenging due to the stochastic nature by which mutations occur in populations and how they are inherited over generations (Hudson, 1990; Wakeley, 2009). Thus, a method for estimating historical patterns of divergences across taxa should explicitly model the stochastic mutational and ancestral processes that generate and filter the genetic variation we observe in present-day genetic data. An appealing approach would be a comparative, Bayesian model-choice method for inferring the probability of competing divergence histories while integrating over uncertainty in mutational and ancestral processes via models of nucleotide substitution and lineage coalescence.

The software package `msBayes` implements such an approach in an approximate-Bayesian model-choice framework (Hickerson et al., 2006; Huang et al., 2011). Recently, however, Oaks et al. (2013) found that `msBayes` can be strongly biased toward spuriously estimating biogeographically interesting models of highly clustered divergence times across taxa. The results of their simulation-based assessment of the method suggest that a combination of broad, uniformly distributed priors on nuisance parameters and a U-shaped prior on divergence models are causing the bias. The model implemented in `msBayes` uses a discrete uniform prior over the number of possible divergence parameters, which Oaks et al. (2013) show imposes a prior on divergence models that puts most of the prior mass on models with either very few or very many divergence times shared across the taxa. Furthermore, uniform priors used for most of the model's nuisance parameters likely hinder the marginal likelihoods of parameter-rich models, i.e., models with many divergence time parameters (Lindley, 1957). Thus, together, the priors on divergence models and nuisance parameters used in `msBayes` could create a strong bias toward models with very few divergence time parameters (Oaks et al., 2013).

Following the recommendations of Oaks et al. (2013), we modify the `msBayes` model to test whether alternative parameterizations and priors will improve the behavior of the approximate-Bayesian model-choice approach. We implement a Dirichlet-process prior over all possible models of divergence, and use gamma and beta probability distributions in place

of uniform priors on many of the model's parameters. Our simulation-based analyses show that the new implementation has improved robustness, accuracy, and power compared to the original model. Also, our results show that both the uniform priors on nuisance parameters and the U-shaped prior on divergence models contribute to the biases of `msBayes`.

3.2 Methods

3.2.1 The model

In this section, we describe our model, which is a modification of the model implemented in `msBayes` (Huang et al., 2011; Oaks et al., 2013). The code implementing the new model is freely available in the open-source software package `dpp-msbayes` (https://github.com/joaks1/dpp-msbayes). Our analyses described below utilize the freely avaliable, open-source software package `PyMsBayes` (https://github.com/joaks1/PyMsBayes), which is a Python API that provides a multi-processing interface to `msBayes` and `dpp-msbayes`. We performed the work described below following the principles of Open Notebook Science. Using version-control software, we make progress in all aspects of the work freely and publicly available in real-time at https://github.com/joaks1/msbayes-experiments. All information necessary to reproduce our results is provided there. We follow much of the notation of Oaks et al. (2013), but modify it to aid in our description of the new model. A summary of our notation can be found in Table 3.1.

We assume an investigator is interested in inferring the distribution of divergence times among Y pairs of populations. For each pair i, n_i genome copies have been sampled, with $n_{1,i}$ copies sampled from population 1, and $n_{2,i}$ sampled from population 2. From these genomes, let k_i be the number of DNA sequence loci collected for population pair i, and K be the total number of unique loci sampled across the Y pairs of populations. We use $X_{i,j}$ to represent the multiple sequence alignment of locus j for population pair i. $\mathbf{X} = (X_{1,1}, \ldots, X_{Y,k_Y})$ is the full dataset, i.e., a vector of sequence alignments for all pairs and

loci. We let $G_{i,j}$ represent the gene tree upon which $X_{i,j}$ evolved according to fixed HKY85 substitution model parameters $\phi_{i,j}$. The investigator must specify the parameters of all $\boldsymbol{\phi} = (\phi_{1,1}, \ldots, \phi_{Y,k_Y})$ substitution models by which the alignments evolved along the $\mathbf{G} = (G_{1,1}, \ldots, G_{Y,k_Y})$ gene trees. Furthermore, the investigator must specify a vector of fixed constants $\boldsymbol{\rho} = (\rho_{1,1}, \ldots, \rho_{Y,k_Y})$ that scale the population-size parameters for known differences in ploidy among loci and/or differences in generation times among population pairs. Lastly, the investigator must also specify a vector of fixed constants $\boldsymbol{\nu} = (\nu_{1,1}, \ldots, \nu_{Y,k_Y})$ that scale the population-size parameters for known differences in mutation rates among loci and/or among taxa.

With $\mathbf{X}$, $\boldsymbol{\phi}$, $\boldsymbol{\rho}$, and $\boldsymbol{\nu}$ in hand, the joint posterior distribution of our model is given by

$$p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta}, \boldsymbol{\upsilon}, \alpha \,|\, \mathbf{X}, \boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\nu}) = \frac{p(\mathbf{X} \,|\, \mathbf{G}, \mathbf{T}, \boldsymbol{\Theta}, \boldsymbol{\upsilon}, \alpha, \boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\nu}) p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta}, \boldsymbol{\upsilon}, \alpha \,|\, \boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\nu})}{p(\mathbf{X})}, \quad (3.1)$$

which can be expanded out into components that are assumed to be independent to get

$$p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta}, \boldsymbol{\upsilon}, \alpha \,|\, \mathbf{X}, \boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\nu}) = \frac{p(\mathbf{X} \,|\, \mathbf{G}, \boldsymbol{\phi}) p(\mathbf{G} \,|\, \mathbf{T}, \boldsymbol{\Theta}, \boldsymbol{\upsilon}, \boldsymbol{\rho}, \boldsymbol{\nu}) p(\boldsymbol{\upsilon} \,|\, \alpha) p(\alpha) p(\mathbf{T}) p(\boldsymbol{\Theta})}{p(\mathbf{X})}, \quad (3.2)$$

where $\mathbf{T} = (T_1, \ldots, T_Y)$ is a vector of population divergence times for each of the Y pairs of populations, $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_Y)$ is a vector of the demographic parameters for each of the Y population pairs, $\boldsymbol{\upsilon} = (\upsilon_1, \ldots \upsilon_K)$ is a vector of locus-specific mutation-rate scaling parameters for each of the K loci, α is the shape parameter of a gamma-distributed prior on υ, and $p(\mathbf{X})$ is the probability of the data (or the marginal likelihood of the model).

To avoid calculating the likelihood terms of Equation 3.2, we distill each sequence alignment X into a vector of insufficient summary statistics S, thus replacing the full dataset $\mathbf{X} = (X_{1,1}, \ldots, X_{Y,k_Y})$ with vectors of summary statistics for each alignment $\mathbf{S}^* = (S^*_{1,1}, \ldots, S^*_{Y,k_Y})$. For each population pair, the means of the summary statistics are calculated across the k loci, reducing the vector to $\mathbf{S}^* = (S^*_1, \ldots, S^*_Y)$. This allows us to estimate the approximate

joint posterior distribution

$$p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta}, \boldsymbol{\upsilon}, \alpha \,|\, B_\epsilon(\mathbf{S}^*), \boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\nu}) = \frac{p(B_\epsilon(\mathbf{S}^*) \,|\, \mathbf{G}, \boldsymbol{\phi}) p(\mathbf{G} \,|\, \mathbf{T}, \boldsymbol{\Theta}, \boldsymbol{\upsilon}, \boldsymbol{\rho}, \boldsymbol{\nu}) p(\boldsymbol{\upsilon} \,|\, \alpha) p(\alpha) p(\mathbf{T}) p(\boldsymbol{\Theta})}{p(B_\epsilon(\mathbf{S}^*))}, \tag{3.3}$$

where $B_\epsilon(\mathbf{S}^*)$ is the multidimensional Euclidean space around the vector of summary statistics, the radius of which is the tolerance ϵ. The sources of approximation are the insufficiency of the statistics and the ϵ being greater than zero. We describe the full model in detail before delving into the numerical method of estimating the approximate model.

Likelihood terms of Equation 3.2

The likelihood terms of Equation 3.2 can be expanded out as a product over population pairs and loci

$$p(\mathbf{X} \,|\, \mathbf{G}, \boldsymbol{\phi}) p(\mathbf{G} \,|\, \mathbf{T}, \boldsymbol{\Theta}, \boldsymbol{\upsilon}, \boldsymbol{\rho}, \boldsymbol{\nu}) = \prod_{i=1}^{Y} \prod_{j=1}^{k_i} p(X_{i,j} \,|\, G_{i,j}, \phi_{i,j}) p(G_{i,j} \,|\, T_i, \Theta_i, \upsilon_j, \rho_{i,j}, \nu_{i,j}). \tag{3.4}$$

The first term, $p(X_{i,j} \,|\, G_{i,j}, \phi_{i,j})$, is the probability of the sequence alignment of locus j for population pair i given the gene tree and HKY85 substitution model parameters (i.e., the "Felsenstein likelihood"; Felsenstein, 1981). The model allows for an intra-locus recombination rate r, which, for simplicity, is assumed to be zero in Equation 3.2. If r is non-zero, this term requires an additional product over the columns (sites) of each sequence alignment to allow sites to have different genealogies. The second term, $\mathrm{p}(G_{i,j} \mid T_i, \Theta_i, \upsilon_j, \rho_{i,j}, \nu_{i,j})$, is the probability of the gene tree under a multi-population coalescent model (i.e., species tree) where the ancestral population of pair i diverges and gives rise to the two sampled descendant populations. Each Θ contains the following demographic parameters: The mutation-scaled sizes of the ancestral, θ_A, and descendant populations, θ_{D1} and θ_{D2}; the proportion of the first, ζ_{D1}, and second population, ζ_{D2}, that persist during bottlenecks that begin immediately after divergence in forward-time; the proportion of time between present and divergence

when the bottlenecks end for both populations, τ_B; and the symmetric migration rate between the descendant populations, m. Thus, the probability of the $n_i - 1$ coalescence times (node heights) of gene tree $G_{i,j}$ is given by a multi-population Kingman-coalescent model (Kingman, 1982) where the ancestral population of size $\theta_{A,i}\rho_{i,j}\nu_{i,j}\upsilon_j$ diverges at time T_i into two descendant populations of constant size $\theta_{D1,i}\rho_{i,j}\nu_{i,j}\upsilon_j\zeta_{D1,i}$ and $\theta_{D2,i}\rho_{i,j}\nu_{i,j}\upsilon_j\zeta_{D2,i}$, which, after time $T_i\tau_{B,i}$, grow exponentially to their present size $\theta_{D1,i}\rho_{i,j}\nu_{i,j}\upsilon_j$ and $\theta_{D2,i}\rho_{i,j}\nu_{i,j}\upsilon_j$, respectively. Following divergence, the descendant populations of pair i exchange migrants at a symmetric rate of m_i.

Prior terms of Equation 3.2

The term $p(\alpha)$ is the prior density function for the shape-parameter of the gamma-distributed prior on rate heterogeneity among loci. This prior is $\alpha \sim U(1, 20)$. The prior probability of the vector of locus-specific mutation-rate scaling parameters given α then expands out as a product over the loci

$$p(\boldsymbol{\upsilon} \,|\, \alpha) = \prod_{j=1}^{K} p(\upsilon_j \,|\, \alpha), \tag{3.5}$$

where each υ is independently and identically distributed (iid) as $\upsilon \sim Gamma(\alpha, 1/\alpha)$. If the recombination rate r is allowed to be non-zero, the prior term $p(r)$ would be added to Equation 3.2, and the prior would be $r \sim Gamma(a_r, b_r)$.

The prior term for the demographic parameters, $p(\boldsymbol{\Theta})$, expands out into its components and as a product over the Y pairs of populations

$$p(\boldsymbol{\Theta}) = \prod_{i=1}^{Y} p(\theta_{A,i})p(\theta_{D1,i})p(\theta_{D2,i})p(\zeta_{D1,i})p(\zeta_{D2,i})p(\tau_{B,i})p(m_i). \tag{3.6}$$

The priors for the demographic parameters are: $\theta_A \sim Gamma(a_{\theta_A}, b_{\theta_A})$, $\theta_{D1} \sim Gamma(a_{\theta_D}, b_{\theta_D})$, $\theta_{D2} \sim Gamma(a_{\theta_D}, b_{\theta_D})$, $\zeta_{D1} \sim Beta(a_{\zeta_D}, b_{\zeta_D})$, $\zeta_{D2} \sim Beta(a_{\zeta_D}, b_{\zeta_D})$, $\tau_B \sim U(0, 1)$, and $m \sim Gamma(a_m, b_m)$.

Priors on divergence models

The prior term for the vector of divergence times for each of the Y pairs of populations can be expanded as

$$p(\mathbf{T}) = p(\mathbf{t})p(\boldsymbol{\tau} \mid \mathbf{t}), \tag{3.7}$$

where $\boldsymbol{\tau}$ is an ordered set of unique divergence time parameters $\{\tau_1, \ldots, \tau_{|\boldsymbol{\tau}|}\}$ whose length $|\boldsymbol{\tau}|$ can range from 1 to Y, and $\mathbf{t}$ is the vector of indices of divergence time parameters $(t_1, \ldots, t_Y)$ that map the divergence times in $\boldsymbol{\tau}$ to each of the Y pairs of populations. Thus, $\mathbf{T}$ is the result of applying the mapping function

$$f(\boldsymbol{\tau}, \mathbf{t}, i) = \tau_{t_i} \tag{3.8}$$

to each population pair i, such that $\mathbf{T} = (T_1 = f(\boldsymbol{\tau}, \mathbf{t}, 1), \ldots, T_Y = f(\boldsymbol{\tau}, \mathbf{t}, Y))$.

Biologically speaking, $\boldsymbol{\tau}$ contains the times of divergence events, the length of which $|\boldsymbol{\tau}|$ is the number of divergence events shared across the Y pairs of populations. For example, if $\boldsymbol{\tau}$ contains a single divergence time parameter τ_1, all Y pairs of populations are constrained to diverge at this time (i.e., $\mathbf{t}$ would contain the index 1 repeated Y times, and $\mathbf{T}$ would contain the value τ_1 repeated Y times), whereas if it contains Y divergence time parameters, the model is fully generalized to allow all of the pairs to diverge at unique times.

Each τ within $\boldsymbol{\tau}$ is *iid* as $\tau \sim Gamma(a_\tau, b_\tau)$. Thus, given the number of unique divergence time classes in $\mathbf{t}$, this determines the probability of prior term $p(\boldsymbol{\tau} \mid \mathbf{t})$. The divergence time parameters are in coalescent units relative to the size of a constant reference population, which we denote θ_C, that is equal to the expectation of the prior on the size of the descendant populations

$$\theta_C = \mathbb{E}(\theta_D). \tag{3.9}$$

Given the size of the descendant populations are *iid* as $\theta_D \sim Gamma(a_{\theta_D}, b_{\theta_D})$, this becomes

$$\theta_C = a_{\theta_D} b_{\theta_D}. \tag{3.10}$$

The τ parameters are in units of θ_C/μ generations, which we denote as $4N_C$ generations. Thus, each τ within $\boldsymbol{\tau}$ is proportional to time and can be converted to the number of generations of the reference population, which we denote τ_{G_C}, by assuming a mutation rate and multiplying by the effective size of the reference population

$$\tau_{G_C} = \tau_{\times} \frac{a_{\theta_D} b_{\theta_D}}{\mu}. \tag{3.11}$$

Thus, for each of the divergence times in $\boldsymbol{\tau}$ to be on the same scale, the relative mutation rates among the pairs of populations are assumed to be fixed and known according to the user-provided values in $\boldsymbol{\nu}$.

As described by Oaks et al. (2013), to get the divergence times in units proportional to the expected number of mutations, we must scale them by the realized population size for locus j of population-pair i

$$\mathcal{T}_{i,j} = T_i \times \frac{\theta_C}{\bar{\theta}_{D,i} \rho_{i,j}}, \tag{3.12}$$

where $\bar{\theta}_{D,i}$ is the mean of θ_{D1} and θ_{D2} for pair i. This gives us the vector of scaled divergence times $\mathcal{T} = (\mathcal{T}_{1,1}, \ldots, \mathcal{T}_{Y,k_Y})$.

As for the prior term $p(\mathbf{t})$, the total sample space of ordered realizations of $\mathbf{t}$ is all the possible partitions of Y elements into 1 to Y categories. The total number of possible partitions is a sum of the Stirling numbers of the second kind over all possible numbers of

categories $|\boldsymbol{\tau}|$

$$B_Y = \sum_{|\boldsymbol{\tau}|=1}^{Y} \left[\frac{1}{|\boldsymbol{\tau}|!} \sum_{j=0}^{|\boldsymbol{\tau}|-1} (-1)^j \binom{|\boldsymbol{\tau}|}{j} (|\boldsymbol{\tau}| - j)^Y \right], \tag{3.13}$$

which is the Bell Number (Bell, 1934). The original `msBayes` model only samples over the unordered realizations of $\mathbf{t}$, such that the sample space is reduced to all the possible integer partitions of Y (Oaks et al., 2013; Huang et al., 2011; Sloan, 2011a,b; Malenfant, 2011). We denote the set of all possible integer partitions of Y as $a(Y)$ and the length of that set as $|a(Y)|$. The advantages, disadvantages, and justification of ignoring the order of $\mathbf{t}$ is discussed in detail below.

We implement two prior probability distributions over the space of all possible discrete divergence models ($\mathbf{t}$). The first simply gives all possible unordered partitions of Y elements equal probability

$$p(\mathbf{t}) = \frac{1}{|a(Y)|}, \tag{3.14}$$

i.e., a discrete uniform prior over all the integer partitions of Y (discrete divergence models). We denote this prior as $\mathbf{t} \sim DU\{a(Y)\}$.

The second prior we implement is based on the Dirichlet-process, which is a stochastic process that groups elements into an unknown number of discrete parameter classes (Ferguson, 1973; Antoniak, 1974). The process is controlled by the concentration parameter χ, which determines how clustered the process will be. We use the Dirichlet process to place a prior over all possible ordered partitions of Y elements, which we denote as $\mathbf{t} \sim DP(\chi)$. We take a hierarchical approach and use a prior probability distribution (i.e., hyperprior) for χ. More specifically, we use a gamma-distributed prior $\chi \sim Gamma(a_\chi, b_\chi)$, where a_χ and b_χ are specified by the user.

This provides a great deal of flexibility for specifying the prior uncertainty regarding

divergence models. The χ determines the prior probability that any two pairs of populations i and j will be assigned to the same divergence time parameter

$$p(t_i = t_j) = \frac{1}{1+\chi}, \tag{3.15}$$

and also the prior probability of the number of divergence time parameters

$$p(|\boldsymbol{\tau}| \mid \chi, Y) = \frac{c(Y, |\boldsymbol{\tau}|)\chi^{|\boldsymbol{\tau}|}}{\prod_{i=1}^{Y}(\chi + i - 1)}, \tag{3.16}$$

where $c(\cdot, \cdot)$ are the unsigned Stirling numbers of the first kind. Equations 3.15 and 3.16 show that smaller values of χ will favor fewer divergence time parameters, and thus more clustered models of divergence, whereas larger values will favor more divergence time parameters, and thus less clustered models of divergence.

3.2.2 Differences between our model and the original `msBayes` model

The prior on divergence models

One of the key differences between our model and that of `msBayes` (Huang et al., 2011) is the prior distribution on divergence models. As discussed in Oaks et al. (2013), in `msBayes` the prior used for $\mathbf{t}$ is a combination of a discrete uniform prior over the possible number of divergence events $|\boldsymbol{\tau}|$ from 1 to Y with a multinomial distribution on the number of times each index of $\boldsymbol{\tau}$ appears in $\mathbf{t}$, with the constraint that all τ parameters are represented at least once (see Equation 2 of Oaks et al. (2013)). We denote this prior used by `msBayes` as $\mathbf{t} \sim DU\{1, \ldots, Y\}$. Oaks et al. (2013) discuss how placing a uniform prior over the number of divergence parameters (denoted $|\boldsymbol{\tau}|$ here, and as Ψ in Huang et al. (2011)) imposes an U-shaped prior over discrete divergence models ($\mathbf{t}$; see Figure 5B of Oaks et al. (2013)). To avoid this, we place priors directly on the sample space of divergence models, thus eliminating the parameter Ψ from the model. We introduce two priors on divergence models: (1) a prior

that is uniform over all unordered divergence models, and (2) a Dirichlet-process prior on all ordered divergence models. The latter provides the user with a great deal of flexibility in expressing their prior beliefs about models of divergence.

Estimating ordered divergence models

As mentioned above, `msBayes` samples over unordered divergence models (i.e., unordered partitions of the Y pairs of populations). That is, the identity of each population pair, and all the information associated with it, is discarded. In our implementation, inference can be done on either unordered or ordered models of divergence. This is discussed in more detail in the description of the ABC implementation below.

The priors on nuisance parameters

Following the recommendations of Oaks et al. (2013), we have replaced the use of continuous uniform distributions for priors on many of the model's parameters (τ, θ_A, θ_{D1}, θ_{D2}, ζ_{D1}, ζ_{D2}, r, m) with more flexible parametric distributions from the exponential family. We introduce gamma-distributed priors for rate parameters that have a sample space of all positive real numbers (τ, θ_A, θ_{D1}, θ_{D2}, r, m), and beta-distributed priors for parameters that are proportions bounded by zero and one (ζ_{D1} and ζ_{D2}). These priors provide the user with much greater flexibility in expressing prior uncertainty regarding the parameters of the model.

Another distinction between our model and that of `msBayes`, is the prior on the sizes of the descendant populations of each pair. As described by Oaks et al. (2013), `msBayes` uses the joint prior

$$\theta_{D1}, \theta_{D2} \sim Beta(1,1) \times 2 \times U(a_\theta, b_{\theta_D}), \tag{3.17}$$

such that the user-specified uniform prior on descendant population size is a prior on

the *mean* size of the two descendant populations of each pair. Under our model, the sizes of the descendant populations of each pair are *iid* as $\theta_{D1} \sim Gamma(a_{\theta_D}, b_{\theta_D})$ and $\theta_{D2} \sim Gamma(a_{\theta_D}, b_{\theta_D})$. This relaxes the assumption that the sizes of the two descendant populations are interdependent and negatively correlated.

Flexibility in parameterizing the model

In our implementation, we provide the ability to control the richness of the model. For the θ parameters, the model can be fully generalized to allow each population pair to have three parameters: θ_A, θ_{D1}, and θ_{D2}. Furthermore, any model of θ parameters nested within this general model can also be specified, including the most restricted model where the ancestral and descendant populations of each pair share a single θ parameter.

We also provide the option of eliminating the parameters associated with the post-divergence bottlenecks in the descendant populations of each pair (τ_B, ζ_{D1}, and ζ_{D2}), which constrains the descendant populations to be of constant size from present back to the divergence event. Also, rather than eliminate the bottleneck parameters, we allow ζ_{D1} and ζ_{D2} to be constrained to be equal, which removes one free parameter from the model for each of the population pairs.

Overall, our implementation allows an investigator to specify a model that has as many as seven parameters per population pair (θ_A, θ_{D1}, θ_{D2}, τ_B, ζ_{D1}, ζ_{D2}, and m) or as few as one parameter per pair (θ), in addition to the $n_i - 1$ coalescence-time parameters.

Time scale

As described above, in our model divergence times are in units of θ_C/μ generations, where θ_C is the expectation of the prior on descendant-population size. As described by Oaks et al. (2013), in `msBayes`, θ_C is half of the upper limit of the continuous uniform prior on the mean of the descendant population sizes.

3.2.3 ABC estimation of the posterior of the model

Sampling from the prior

To estimate the approximate posterior of Equation 3.3, we use an ABC rejection algorithm. The first step of this algorithm entails collecting a random sample of parameter values from the joint prior and their associated summary statistics. Each sample is generated by (1) drawing values of all the model's parameters, which we denote Λ, from their respective prior distributions; (2) rescaling the divergence times $\mathbf{T} = (T_1, \ldots, T_Y)$ from units proportional to time to units proportional to the expected number of mutations via Equation 3.12 to get $\mathcal{T} = (\mathcal{T}_{1,1}, \ldots, \mathcal{T}_{Y,k_Y})$; (3) simulating gene trees $\mathbf{G} = (G_{1,1}, \ldots, G_{Y,k_Y})$ for each locus of each population pair by drawing coalescent times from a multi-population Kingman-coalescent model given the demographic parameters; (4) simulating sequence alignments $\mathbf{X} = (X_{1,1}, \ldots, X_{Y,k_Y})$ along the gene trees under the HKY85 substitution parameters $\boldsymbol{\phi} = (\phi_{1,1}, \ldots, \phi_{Y,k_Y})$ that have the same number of sequences and sequence lengths as the observed dataset; (5) calculating population genetic summary statistics $\mathbf{S} = (S_{1,1}, \ldots, S_{Y,k_Y})$ from the simulated sequence alignments; and (6) reducing the summary statistics to the means across loci for each population pair to get $\mathbf{S} = (S_1, \ldots, S_Y)$, which is the same summary statistic vector estimated from the observed data $\mathbf{S}^*$. After repeating this procedure $\mathbf{n}$ times, we obtain a random sample of parameter vectors $\boldsymbol{\Lambda} = (\Lambda_1, \ldots, \Lambda_{\mathbf{n}})$ from the model prior and their associated vectors of summary statistics $\mathbb{S} = (\mathbf{S_1}, \ldots, \mathbf{S_n})$.

For all of our analyses below, we use four summary statistics for each of the pairs of populations: π (Tajima, 1983), θ_W (Watterson, 1975), π_{net} (Takahata and Nei, 1985), and $SD(\pi - \theta_W)$ (Tajima, 1989). Furthermore, in addition to model parameters, each sample Λ also contains four statistics that summarize $\mathbf{T}$: the mean ($\bar{T}$), variance (s_T^2), dispersion index ($D_T = s_T^2/\bar{T}$), and the number of divergence time parameters ($|\boldsymbol{\tau}|$). These have been denoted as $E(\tau)$, $Var(\tau)$, Ω, and Ψ, respectively (Hickerson et al., 2006; Huang et al., 2011; Oaks et al., 2013). We use the new notation to avoid confusion that the values represent

prior or posterior expectations, and to indicate that Ω is not a parameter of the model and Ψ is no longer a parameter in the new implementation.

Ordering of taxon-specific summary statistics

As alluded to in the model description, `msBayes` does not maintain the order of the taxon-specific summary statistics S within each $\mathbf{S}$. Rather, the summary statistics are re-ordered by descending values of average pairwise differences between the descendant populations (π_b; Nei and Li, 1979; Huang et al., 2011). This has the advantage of reducing the sample space of the number of discrete divergence models $\mathbf{T}$, but there are at least two disadvantages. First, additional information in the data is lost. By discarding the identity of the Y pairs of populations, all pair-specific information about the amount of data (e.g., the number gene copies collected from each of the populations [n_1 and n_2], the number of loci, and the length of the loci), and the taxon- and locus-specific parameters (ϕ, ν, ρ, and υ) is lost. Second, the results are more difficult to interpret, because they can be no longer be directly associated to the taxa under study.

The original descriptions of the `msBayes` model claim that this re-ordering is justified by the fact that of π_b (and other summary statistics) are unrelated to the sample sizes n_1 and n_2 of each pair and are thus exchangeable (Hickerson et al., 2006; Huang et al., 2011). This is actually incorrect for two of reasons. First, the expectation of π_b is not independent of samples sizes. If there are more than one coalescent events in the ancestor, which is expected to be common on phylogeographic timescales, more samples will increase the probability of capturing these deeper events, and thus affect the average pairwise differences between the descendant populations. Also, other statistics that estimate gross diversity (e.g., π and θ_W) are clearly not independent of sample size. Second, for variables to be exchangeable, they do not need to be independent, but their marginal distributions must be the same (i.e., they must be identically distributed). The simulated alignments and their summary statistics are not identically distributed because of differing sample sizes *and* taxon- and locus-specific

parameters ϕ, ν, and ρ.

Thus, the theoretical basis for this reshuffling is questionable. It may be justifiable as an additional coarsening of the data, by removing the information associated with the identity of the population pairs. However, we do not provide a formal proof that this re-ordering of the pairs does not introduce bias. One can imagine situations in which the sampling intensity (i.e., the number of gene copies, loci, and locus length) is highly skewed across pairs and/or there is large heterogeneity in mutational processes (e.g., ϕ, ν) among the pairs. In such cases, it seems possible that discarding this information could cause bias. Furthermore, given that (1) part of the motivation for re-ordering by π_b is to minimize Euclidean distances between simulated datasets in which the true model has a single divergence event, and (2) the reordering increases the model's tendency to infer a single divergence event (Huang et al., 2011), it seems possible that this approach could be biasing the method (Oaks et al., 2013).

To maintain compatibility and comparability with `msBayes` we maintain the re-ordering of taxon-specific summary statistics by π_b. However, we also allow the order to be maintained, and ordered divergence models to be estimated.

Obtaining an approximate posterior from the prior samples

We use a rejection algorithm to retain an approximate posterior sample of Λ from the prior sample $\mathbf{\Lambda} = (\Lambda_1, \ldots, \Lambda_\mathbf{n})$. First, the observed summary statistics $\mathbf{S}^*$, and the summary statistics of the prior samples $\mathbb{S} = (\mathbf{S_1}, \ldots, \mathbf{S_n})$, are standardized using the means and standard deviations of the statistics from the prior sample (i.e., the prior mean is subtracted from each statistic, and the difference is divided by the prior standard deviation). After all statistics are standardized, the Euclidean distance between $\mathbf{S}^*$ and each $\mathbf{S}$ within $\mathbb{S}$ is calculated. The samples that fall within a range of tolerance ϵ around $\mathbf{S}^*$ are retained. The range of tolerance is determined by specifying the number of desired posterior samples to be retained. Post-hoc adjustment of the posterior sample can also be performed with a number of regression techniques (Beaumont et al., 2002; Blum and François, 2009; Leuen-

berger and Wegmann, 2010). For our analyses, we use the general linear model (GLM) regression adjustment (Leuenberger and Wegmann, 2010) as implemented in `ABCtoolbox` (v1.1; Wegmann et al., 2010), which Oaks et al. (2013) showed performed very similarly to weighted local-linear regression and multinomial logistic regression adjustments (Beaumont et al., 2002) for `msBayes` posteriors.

3.2.4 Assessing model-choice behavior and robustness

Following the simulation-based approach of Oaks et al. (2013), we characterize the behavior of several models under the ideal conditions where the data are generated from parameters drawn from the same prior distributions used for analysis (i.e., the prior is correct). We selected four model priors for these analyses (Table 3.2): (1) A model to represent the original `msBayes` model, $M_{msBayes}$, with priors $\mathbf{t} \sim DU\{1,\ldots,Y\}$, $\tau \sim U(0,10)$, $\theta_A \sim U(0,0.05)$, and $\bar{\theta}_D \sim U(0,0.05)$; (2) a Dirichlet-process prior model, M_{DPP}, with priors $\mathbf{t} \sim DP(\chi \sim Gamma(2,2))$, $\tau \sim Exp(0.3464)$, $\theta_A \sim Exp(40)$, and $\theta_D \sim Exp(40)$; (3) a model with uniform prior probability over unordered divergence models, $M_{Uniform}$, with priors $\mathbf{t} \sim DU\{a(Y)\}$, $\tau \sim Exp(0.3464)$, $\theta_A \sim Exp(40)$, and $\theta_D \sim Exp(40)$; and (4) a model with the `msBayes` prior on unordered divergence models, but with exponential priors on nuisance parameters, $M_{Ushaped}$, with priors $\mathbf{t} \sim DU\{1,\ldots,Y\}$, $\tau \sim Exp(0.3464)$, $\theta_A \sim Exp(40)$, and $\theta_D \sim Exp(40)$. We selected the exponential prior on divergence time used in models M_{DPP}, $M_{Uniform}$, and $M_{Ushaped}$ to have the same variance as the uniform prior in model $M_{msBayes}$. We selected the exponential prior on population size used in models M_{DPP}, $M_{Uniform}$, and $M_{Uniform}$ to have the same mean as the uniform prior in model $M_{msBayes}$, so that all four models have the same θ_C and thus the same units of time. All of the models were the same in other respects, with three free θ parameters for each population pair, two uniformly distributed ($beta(1,1)$) ζ_D parameters per pair, no migration, no recombination, and unordered divergence models. For all of our simulations, we used a simulated data structure of eight population pairs, with a single 1000 base-pair locus sampled from 10

individuals from each population.

For each of the four models, we generated 1×10^6 samples from the prior, and simulated 50,000 pseudo-replicate datasets, also drawn from the prior. We then analyzed each of the replicate datasets, retaining a posterior of 1000 samples from the respective prior. A GLM-regression adjusted posterior was also estimated from each of the posterior samples (Leuenberger and Wegmann, 2010). To assess the robustness of each of the four models, we also analyzed the pseudo-replicate datasets simulated under the other three models. Overall, for each model, we produced 200,000 posterior estimates, 50,000 from the replicate datasets simulated under that model, and 150,000 estimated from the replicate datasets simulated under the other three models.

For each set of 50,000 simulated datasets, we used the posterior estimates to assess the model-choice behavior of each model. We did this by assigning the 50,000 estimates of the posterior probability of one divergence event to 20 bins of width 0.05, and plotted the estimated posterior probability of each bin against the proportion of replicates in that bin with a true value consistent with one divergence event (Huelsenbeck and Rannala, 2004b; Oaks et al., 2013). We did this using two criteria for the one divergence model: (1) the number of divergence time parameters ($|\boldsymbol{\tau}| = 1$) and (2) the dispersion index of divergence times ($D_T < 0.01$ Hickerson et al., 2006; Huang et al., 2011). We used the one-divergence model to assess model-choice behavior, because it is often of biogeographic interest and is easily comparable among the three different priors used on divergence models.

In addition to the four models above, we also assessed the behavior of a model with ordered divergence models (i.e., the order of the taxon-specific summary statistic vectors were maintained for the observed and simulated datasets). For this, we used a model with identical priors as M_{DPP}, but that samples over ordered divergence models. We denote this model as M°_{DPP}. We generated 1×10^6 prior samples and 50,000 pseudo-replicate datasets, and analyzed them as above. We were not able to analyze the replicate datasets of the other models under the ordered model, because the identity of the population pairs is not

contained in the simulations of the other models.

3.2.5 Assessing power

We evaluated the power of the same four models (Table 3.2) to detect random variation in divergence times using methods similar to Oaks et al. (2013). The only difference from the prior models used in the validation analyses above is the prior on the concentration parameter χ for the M_{DPP} model. Rather than the prior $\mathbf{t} \sim DP(\chi \sim Gamma(2, 2))$, we use the prior $\mathbf{t} \sim DP(\chi \sim Gamma(1.5, 18.1))$ over the discrete divergence models for the model M_{DPP}. The reason for this difference is because for all of our power simulations, we used a simulated data structure identical to that of the empirical dataset of Philippine vertebrates analyzed by Oaks et al. (2013), which consists of 22 pairs of populations. For each of the four models, we generated 2×10^6 samples from the prior. We then generated pseudo-observed datasets from three series of models. One series of models, which we denote $\mathcal{M}_{msBayes}$, is identically distributed as $M_{msBayes}$ except that the divergence times for each of the 22 pairs of populations are randomly drawn from a series of uniform distributions, $U(0, \tau_{max})$, where τ_{max} was set to: 0.2, 0.4, 0.6, 0.8, 1.0, and 2.0, in $4N_C$ generations. A second series of models, $\mathcal{M}_{Uniform}$, are identically distributed as $M_{Uniform}$ except that the 22 divergence times are randomly drawn from the same series of uniform priors as above. The third series of models, $\mathcal{M}_{Exp}$, is also identically distributed as $M_{Uniform}$ except that the 22 divergence times are randomly drawn from a series of of exponential distributions: $Exp(17.3205)$, $Exp(8.6603)$, $Exp(5.7735)$, $Exp(4.3301)$, $Exp(3.4641)$, and $Exp(1.7321)$. These exponential distributions have the same variance as their uniform counterparts in the first two series of models.

For each of the six models in each of the three series of models, we simulated 1000 pseudo-replicate datasets (18,000 datasets in total). We then analyzed each replicate dataset under all four prior models (Table 3.2), producing 72,000 posterior estimates, each with 1000 samples. We also estimated a GLM-regression adjusted posterior from each of the posterior samples (Leuenberger and Wegmann, 2010).

3.3 Results

3.3.1 Validation analyses: Estimation accuracy

In terms of estimating summary statistics of model parameters ($|\boldsymbol{\tau}|$, D_T, and $\bar{T}$), the M_{DPP} performs similarly when applied to datasets generated under all four models of Table 3.2 (Figs. 3.1, 3.5, 3.9 and 3.13). The same is true for models $M_{Uniform}$ (Figs. 3.2, 3.6, 3.10 and 3.14) and $M_{Ushaped}$ (Figs. 3.4, 3.8, 3.12 and 3.16). All three of the models with exponential priors on nuisance parameters (M_{DPP}, $M_{Uniform}$, and $M_{Ushaped}$) perform well when applied to the data generated under the original `msBayes` model, $M_{msBayes}$ (Figs. 3.9, 3.10 and 3.12). The $M_{msBayes}$ model also performed well on its own datasets (Figure 3.11), but performed poorly when applied to the data generated under the other three models (Figs. 3.3, 3.7 and 3.15). Thus, the newly implemented models display greater robustness than the original model.

Posterior adjustment with GLM regression does improve the estimation accuracy for the continuous variables D_T and $\bar{T}$, however, this is not always true for estimating the discrete variable $|\boldsymbol{\tau}|$ (Figs. 3.1–3.16).

3.3.2 Validation analyses: Model-choice accuracy

The `msBayes` model, and our modification of it, is a model-choice method with the primary purpose of estimating the probabilities of models of divergence across taxa. Thus, it is critical to assess the method's ability to accurately estimate the posterior probabilities of divergence models. Our results demonstrate that the unadjusted estimates of divergence models are generally more accurate than the regression-adjusted estimates, especially when using $D_T < 0.01$ to assess the probability of the one divergence model (Figs. 3.17–3.32). Thus, we will focus our discussion of the results on the unadjusted estimates.

The model-choice behavior of the M_{DPP} is relatively accurate when applied to datasets generated under models M_{DPP}, $M_{Uniform}$, and $M_{msBayes}$ (Figs. 3.17, 3.21 and 3.25). It

is less accurate and tends to underestimate the probability of the one-divergence model when applied to the data generated under $M_{Ushaped}$ (Figure 3.29). The $M_{Uniform}$ model performs almost identically as M_{DPP} in terms of model-choice accuracy (Figs. 3.18, 3.22, 3.26 and 3.30).

The original `msBayes` model $M_{msBayes}$ performs reasonably well when applied to its own datasets (Figure 3.27), but it performs poorly when applied to data generated under any of the other models (Figs. 3.19, 3.23 and 3.31). In these cases the $M_{msBayes}$ model is strongly biased towards over-estimating the posterior probability of the one-divergence model, especially when the divergence models are not distributed under its U-shaped prior (Figs. 3.19 and 3.23). The other model with the U-shaped prior on divergence models, but exponential priors on nuisance parameters ($M_{Ushaped}$), performs similarly to $M_{msBayes}$ in that it performs well when applied to its own data, but overestimates the probability of the one-divergence model when applied to data generated by the other models (Figs. 3.20, 3.24, 3.32 and 3.32). However, the bias toward over-estimating the probability of the one-divergence model is stronger in the $M_{msBayes}$ model than $M_{Ushaped}$.

Overall, the result suggest that the M_{DPP} and $M_{Uniform}$ models are relatively robust in terms of model-choice accuracy, and when model violations do cause them to be biased, they tend to under-estimate the probability of the biogeographically interesting model of a single divergence event. In contrast, the $M_{msBayes}$ model is very sensitive to model violations, and strongly over-estimates the one-divergence model whenever the model is misspecified. Furthermore, the results suggest that using exponentially distributed priors on nuisance parameters rather than uniform priors helps the $M_{Ushaped}$ model perform better than $M_{msBayes}$, but it is still hindered by the U-shaped prior on divergence models and tends to over-estimate the probability of the one-divergence model whenever there are violations of the model.

3.3.3 Validation analyses: Ordered divergence models

Our results show that the ordered models of divergence, M°_{DPP}, perform similarly to sampling over unordered models (M_{DPP}) when the prior is correct (Figs. 3.33 and 3.34). This suggests that the method is not adversely affected by the large increase in the number of possible discrete models (from 22 unordered to 4140 ordered models) when there are eight pairs of populations.

3.3.4 Power analyses: Estimation accuracy

All of the models we evaluated (Table 3.2) struggle to estimate the dispersion index of divergence times D_T regardless of which of the three series of models (Table 3.3) the data were generated under (Figs. 3.35–3.46). When the divergence times of the 22 population pairs are randomly drawn from a series of exponential priors ($\mathcal{M}_{Exp}$), the M_{DPP} model is the best estimator of D_T, followed by $M_{Uniform}$ (figs. 3.35 and 3.36). The $M_{msBayes}$ model is strongly biased toward underestimating D_T, estimating values of zero for most of the replicates across all the data models of $\mathcal{M}_{Exp}$ (Figure 3.37). The results of the $M_{Ushaped}$ model are intermediate between those of $M_{msBayes}$ and the new models M_{DPP} and $M_{Uniform}$ (Figure 3.38).

When the true divergence times are randomly drawn from a series of uniform priors ($\mathcal{M}_{Uniform}$), all of the models poorly estimate D_T (Figs. 3.39–3.42). The M_{DPP} and $M_{Uniform}$ models tend to over-estimate the variance in divergence times (Figs. 3.39 and 3.40), whereas the $M_{msBayes}$ model, again, underestimates D_T, estimating values of zero for most replicates across all the data models of $\mathcal{M}_{Uniform}$ (Figure 3.41). Again, the performance of the $M_{Ushaped}$ model is intermediate between the $M_{msBayes}$ and $M_{DPP}/M_{Uniform}$ models (Figure 3.42). This pattern of results across the four models is very similar to their behavior when applied to data simulated under the series of models of $\mathcal{M}_{msBayes}$ (Figs. 3.43–3.46).

3.3.5 Power analyses: Model choice

There is a large amount of heterogeneity in estimating the number of divergence events ($|\boldsymbol{\tau}|$) both among the prior models, and across the different series of data models (Table 3.3) (Figs. 3.47–3.58). The M_{DPP} model tends to infer highly clustered divergence models across all three series of data models when the divergences are recent, but performs the best of the four prior models when divergences are random over a timescale of about one to two coalescent units (Figs. 3.47, 3.51 and 3.55). The $M_{Uniform}$ model performs the best when divergences are random over the most recent timescales, but tends to only infer a tight cluster of $|\boldsymbol{\tau}|$ values even when divergences are random over longer periods (Figs. 3.48, 3.52 and 3.56). The $M_{msBayes}$ model performs the worst of the four prior models across all three series of data models, inferring a single divergence event across most of the 18,000 simulations (Figs. 3.49, 3.53 and 3.57). Even when the simulated data are identically distributed as the $M_{msBayes}$ model (except for the divergence times) the $M_{msBayes}$ is still strongly biased towards a single event (Figure 3.57), whereas the M_{DPP} model is the least likely to infer a single event when applied to these data (Figure 3.55). Using exponential priors on nuisance parameters does increase the power of the $M_{Ushaped}$ model compared to $M_{msBayes}$ across all three series of data models, but the U-shaped prior still prevents the model from performing as well as the M_{DPP} and $M_{Uniform}$ models (Figs. 3.50, 3.54 and 3.58).

The improved power of the new models are even more pronounced when looking at estimates of the dispersion index of divergence times (D_T) across the simulations (Figs. 3.59–3.70). The M_{DPP} and $M_{Uniform}$ models perform similarly across all three series of data models, inferring values of D_T consistent with one divergence event ($D_T < 0.01$) in almost none of the replicates across all the simulations (Figs. 3.59, 3.60, 3.63, 3.64, 3.67 and 3.68). The $M_{msBayes}$ model infers values consistent with a single divergence event in most of the replicates across all the simulations (Figs. 3.61, 3.65 and 3.69). Using exponential priors on nuisance parameters greatly increases the power of the $M_{Ushaped}$ model relative to $M_{msBayes}$, but it still has less power than the models with DPP or uniform priors across divergence

models (Figs. 3.62, 3.66 and 3.70).

When looking at the estimated posterior probability of the single-divergence model across the power analyses, we also see increased power of the new models (Figs. 3.71–3.82). The M_{DPP} model estimates low posterior probability of $|\boldsymbol{\tau}| = 1$ across all of the simulations (Figs. 3.71, 3.75 and 3.79). The M_{DPP} does infer strong support for one divergence as gauged by a Bayes factor of greater than 10. However, this is not that unexpected given the low prior probability of a single divergence under this model. Also, the Bayes factor calculations for the M_{DPP} replicates are only approximate; they are calculated based on the prior mean of the concentration parameter χ.

The $M_{Uniform}$ model infers very low posterior probabilities for the one-divergence model across all the simulations (Figs. 3.72, 3.76 and 3.80), whereas the $M_{msBayes}$ model infers high posterior probabilities of a single divergence for most replicates across all simulations (Figs. 3.73, 3.77 and 3.81). Again, we see that the exponential priors on nuisance parameters greatly increase power and result in lower estimates of the probability of one divergence when comparing $M_{Ushaped}$ to $M_{msBayes}$ (Figs. 3.74, 3.78 and 3.82).

Lastly, when we look at the estimated posterior probability of D_T being consistent with one divergence parameter ($p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$), we see the same pattern of model behavior, with M_{DPP} and $M_{Uniform}$ inferring low probabilities across all simulations, $M_{msBayes}$ inferring high probabilities, and $M_{Ushaped}$ inferring intermediate values (Figs. 3.83–3.94).

3.4 Discussion

Our results demonstrate that using alternative priors on nuisance parameters and divergence models improved the behavior of the `msBayes` model. In our implementation, model-choice estimation is more accurate and shows greater robustness to model violations (Figs. 3.17–3.32). The original model is very sensitive to model violations and, when present, strongly over-estimates the probability of one divergence event shared across all taxa (Figs. 3.19, 3.23

and 3.31). When more appropriate priors are used for nuisance parameters, and either a Dirichlet-process or uniform prior applied across divergence models, the model is less sensitive to violations (Figs. 3.18, 3.21, 3.25, 3.26, 3.29 and 3.30), and, when they do cause bias, the method is conservative and tends to under-estimate the probability of biogeographically interesting models of temporally clustered divergences (Figs. 3.29 and 3.30).

Furthermore, the modifications to the model also improve the method's power to detect random variation in divergence times, reducing the bias toward estimating clustered divergences (Figs. 3.47–3.94). Our results are similar to those of Oaks et al. (2013) in that we also find the `msBayes` model to infer strong support for clustered divergences when divergences are random over quite broad timescales (Figs. 3.73, 3.77, 3.81, 3.85, 3.89 and 3.93). Our results expand on this by showing that this behavior is consistent across a range conditions underlying the data. When using the new priors, the method is less prone to spurious inference of clustered divergence models, and much less likely to incorrectly infer such models with strong support.

By evaluating a model intermediate between the old and new implementation ($M_{Ushaped}$), we were able to determine the relative effects of the modifications to model. Across all of our analyses, the results show that using better priors on nuisance parameters alone does improve the performance of the method. The magnitude of the bias toward inferring support for the one-divergence model when there are model violations is reduced when the exponential priors are used in place of the uniform priors (Figs. 3.19, 3.20, 3.23, 3.24, 3.28 and 3.31). Furthermore, better priors on nuisance parameters improves the method's power to detect temporally random divergences. Throughout our analyses, the intermediate model ($M_{Ushaped}$) performs better than the `msBayes` model, but not as well as the models with alternative priors on divergence models. This suggests, as predicted by Oaks et al. (2013), that the bias of `msBayes` toward models of temporally clustered divergences is caused by a combination of (1) uniform priors on nuisance parameters hindering the marginal likelihoods of models with more τ parameters and (2) the U-shaped prior on divergence models giving

low prior density to models with intermediate numbers of divergence parameters.

While our modifications improve the behavior of the model, we urge caution when using the method and interpreting its results. The method attempts to approximate the posterior of a very parameter rich model using relatively little information from the data. For example, when applied to the dataset of Philippine vertebrates analyzed by Oaks et al. (2013) the model can have as many as 604–625 free parameters (depending on $|\boldsymbol{\tau}|$), and samples over 1002 divergence models. Our modifications allow this model to be simplified to 471–492 free parameters. Sampling over hundreds of parameters and more than a thousand models is a difficult inference problem, especially when only using a small number of summary statistics calculated from the sequence alignments of each taxon pair. The population-genetic summary statistics used by our method contain little information about many of the free parameters in the model. Thus, we expect the improved method will still be sensitive to priors, and the power, while improved, may still be low. Furthermore, ABC methods in general have been shown to be biased (Robert et al., 2011) for model choice. As recommended by Oaks et al. (2013), any results from the method should be accompanied by assessments of prior sensitivity and power. We encourage investigators to view this method and its predecessor as a means of exploring their data rather than a rigorous means of evaluating hypotheses.

Given the difficulty of this estimation problem, we anticipate that full-likelihood methods that can leverage all of the information present in the sequence data will become increasingly important for robustly estimating the temporal history of divergences across taxa (Sukumaran, 2012). With improving numerical methods for sampling over models of differing dimensionality (Green, 1995; Lemey et al., 2009) and increasing efficiency of likelihood calculations (Ayres et al., 2012), analyzing rich comparative phylogeograpical models in a full-likelihood Bayesian framework is becoming computationally practical, especially when considering that the computational load of simulating millions of random datasets from the prior under the ABC approach is nontrivial.

3.5 Tables

Table 3.1. Summary of the notation used throughout Chapter 3; modified from Oaks et al. (2013).

Symbol	Description
Y	Number of population pairs.
n_i	The number of genome copies sampled from population pair i, with $n_{1,i}$ sampled from population 1, and $n_{2,i}$ from population 2.
k_i	Number of loci sampled from population pair i.
K	Total number of unique loci sampled.
$X_{i,j}$	Sequence alignment of locus j sampled from population pair i.
$S^*_{i,j}$	Population genetic summary statistics calculated from $X_{i,j}$.
$\mathbf{X}$	Vector containing the sequence alignments of each locus from each population pair: $(X_{1,1}, \ldots, X_{Y,k_Y})$.
$\mathbf{S}^*$	Vector containing the summary statistics of each locus from each population pair: $(S^*_{1,1}, \ldots, S^*_{Y,k_Y})$.
$B_\epsilon(\mathbf{S}^*)$	Multi-dimensional Euclidean space around the observed summary statistics, $\mathbf{S}^*$.
ϵ	Radius of $B_\epsilon(\mathbf{S}^*)$, i.e., the tolerance of the ABC estimation.
$G_{i,j}$	Gene tree of the sequences in $X_{i,j}$.
$\mathbf{G}$	Vector containing the gene trees of each locus from each population pair: $(G_{1,1}, \ldots, G_{Y,k_Y})$.
$\lvert\boldsymbol{\tau}\rvert$	Number of unique population divergence times among the Y population pairs.
τ	One of the $\lvert\boldsymbol{\tau}\rvert$ *iid* divergence time paramters in $4N_C$ generations.
$\boldsymbol{\tau}$	Ordered set of divergence time parameters: $\{\tau_1, \ldots, \tau_{\lvert\boldsymbol{\tau}\rvert}\}$.
t_i	The index of the divergence time parameter in $\boldsymbol{\tau}$ that is mapped to population pair i.
$\mathbf{t}$	Vector of divergence time parameter indices: $(t_1, \ldots, t_Y)$.
T_i	Time of divergence in $4N_C$ generations between the populations of pair i.
$\mathbf{T}$	Vector of divergence times for each of the population pairs: $(T_1, \ldots, T_Y)$.
$\mathcal{T}_{i,j}$	Scaled time of divergence between the populations of pair i for locus j.
$\boldsymbol{\mathcal{T}}$	Vector containing the scaled divergence times of each locus from each population pair: $(\mathcal{T}_{1,1}, \ldots, \mathcal{T}_{Y,k_Y})$.
$\theta_{D1,i}$, $\theta_{D2,i}$	θ of the 1^{st} and 2^{nd} descendent population, respectively, of pair i.
$\theta_{A,i}$	θ of the population ancestral to pair i.
$\boldsymbol{\theta_{D1}}$, $\boldsymbol{\theta_{D2}}$	Vectors $(\theta_{D1,1}, \ldots, \theta_{D1,Y})$ and $(\theta_{D2,1}, \ldots, \theta_{D2,Y})$, respectively.
$\boldsymbol{\theta_A}$	Vector containing the θ_A parameters for each population pair: $(\theta_{A,1}, \ldots, \theta_{A,Y})$.
υ_j	mutation-rate scaling parameter for locus j.
$\boldsymbol{\upsilon}$	Vector containing the locus-specific mutation-rate parameters: $(\upsilon_1, \ldots, \upsilon_K)$.
α	Hyper-parameter for the shape of the gamma prior distribution on each υ.
$\zeta_{D1,i}$, $\zeta_{D2,i}$	θ-scaling parameters that determine the magnitude of the population bottleneck in the 1^{st} and 2^{nd} descendant population of pair i, respectively. The bottleneck in each descendant population begins immediately after divergence.
$\boldsymbol{\zeta_{D1}}$, $\boldsymbol{\zeta_{D2}}$	Vectors $(\zeta_{D1,1}, \ldots, \zeta_{D1,Y})$ and $(\zeta_{D2,1}, \ldots, \zeta_{D2,Y})$, respectively.
$\tau_{B,i}$	Proportion of time between present and τ_i when the bottleneck ends for the descendant populations of pair i.
$\boldsymbol{\tau_B}$	Vector containing the τ_B parameters for each population pair: $(\tau_{B,1}, \ldots, \tau_{B,Y})$.
m_i	Symmetric migration rate between the descendant populations of pair i.
$\mathbf{m}$	Vector containing the migration rates for each population pair: $(m_1, \ldots, m_Y)$.
$\rho_{i,j}$	θ-scaling constant provided by the user for locus j of pair i. This constant is intended to allow the user to scale θ for differences in ploidy among loci or differences in generation times among taxa.
$\nu_{i,j}$	θ-scaling constant provided by the user for locus j of pair i. This constant is intended to allow the user to scale θ for differences in mutation rates among loci or among taxa.
$\boldsymbol{\rho}$	Vector of ploidy and/or generation-time scaling constants: $(\rho_{1,1}, \ldots, \rho_{Y,k_Y})$.
$\boldsymbol{\nu}$	Vector of mutation-rate scaling constants: $(\nu_{1,1}, \ldots, \nu_{Y,k_Y})$.
$\bar{T}$	Mean of $\mathbf{T}$.
s^2_T	Variance of $\mathbf{T}$.
D_T	Dispersion index of $\mathbf{T}$ ($s^2_T/\bar{T}$).
$\mathbf{n}$	Number of samples from the joint prior.
Λ	Vector of parameter values drawn from the joint prior.
$\mathbf{S}$	Vector containing the summary statistics calculated from data simulated under parameter values drawn from the prior (Λ).
$\boldsymbol{\Lambda}$	Random sample of $\Lambda_1, \ldots, \Lambda_\mathbf{n}$ drawn form the prior.
$\mathbb{S}$	Summary statistic vectors $\mathbf{S_1}, \ldots, \mathbf{S_n}$ for each $\Lambda_1, \ldots, \Lambda_\mathbf{n}$ drawn from the prior.

Table 3.2. The model priors evaluated in our simulation-based analyses. For the M_{DPP} model, the prior on the concentration parameter, $\chi \sim Gamma(\cdot, \cdot)$, was set to $Gamma(2, 2)$ for the validation analyses and $Gamma(1.5, 18.1)$ for the power analyses.

	Priors			
Model	$\mathbf{t}$	τ	θ_A	θ_D
$M_{msBayes}$	$\mathbf{t} \sim DU\{1, \ldots, Y\}$	$\tau \sim U(0, 10)$	$\theta_A \sim U(0, 0.05)$	$\bar{\theta}_D \sim U(0, 0.05)$
M_{DPP}	$\mathbf{t} \sim DP(\chi \sim Gamma(\cdot, \cdot))$	$\tau \sim Exp(0.3464)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
$M_{Uniform}$	$\mathbf{t} \sim DU\{a(Y)\}$	$\tau \sim Exp(0.3464)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
$M_{Ushaped}$	$\mathbf{t} \sim DU\{1, \ldots, Y\}$	$\tau \sim Exp(0.3464)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$

Table 3.3. The models we used to simulate pseudo-replicate datasets for assessing the power of the models in Table 3.2.

	Priors			
Model series	$\mathbf{t}$	τ	θ_A	θ_D
$\mathcal{M}_{msBayes}$	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim U(0, 0.2)$	$\theta_A \sim U(0, 0.05)$	$\bar{\theta}_D \sim U(0, 0.05)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim U(0, 0.4)$	$\theta_A \sim U(0, 0.05)$	$\bar{\theta}_D \sim U(0, 0.05)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim U(0, 0.6)$	$\theta_A \sim U(0, 0.05)$	$\bar{\theta}_D \sim U(0, 0.05)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim U(0, 0.8)$	$\theta_A \sim U(0, 0.05)$	$\bar{\theta}_D \sim U(0, 0.05)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim U(0, 1.0)$	$\theta_A \sim U(0, 0.05)$	$\bar{\theta}_D \sim U(0, 0.05)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim U(0, 2.0)$	$\theta_A \sim U(0, 0.05)$	$\bar{\theta}_D \sim U(0, 0.05)$
$\mathcal{M}_{Uniform}$	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim U(0, 0.2)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim U(0, 0.4)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim U(0, 0.6)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim U(0, 0.8)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim U(0, 1.0)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim U(0, 2.0)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
$\mathcal{M}_{Exp}$	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim Exp(17.3205)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim Exp(8.6603)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim Exp(5.7735)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim Exp(4.3301)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim Exp(3.4641)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$
	$\lvert\boldsymbol{\tau}\rvert = 22$	$\tau \sim Exp(1.7321)$	$\theta_A \sim Exp(40)$	$\theta_D \sim Exp(40)$

3.6 Figures

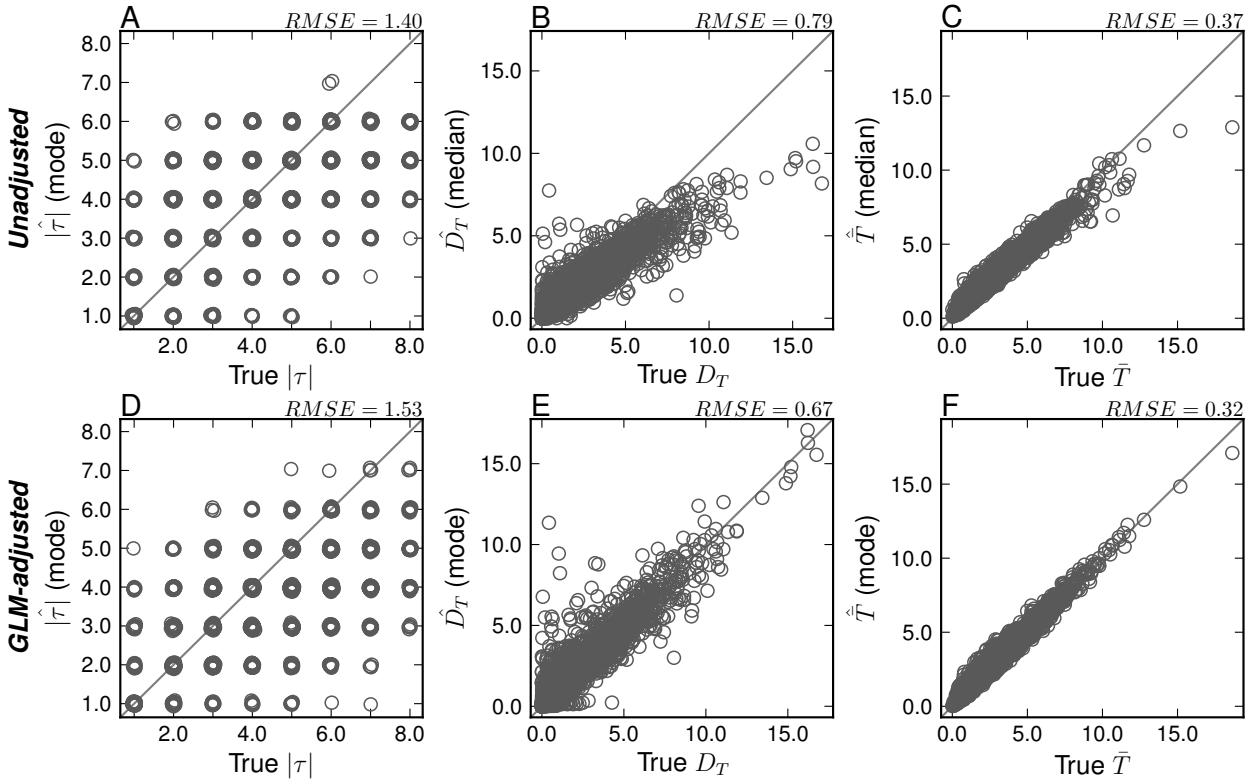

Figure 3.1. Estimation accuracy for model M_{DPP} when analyzing data generated under M_{DPP}. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

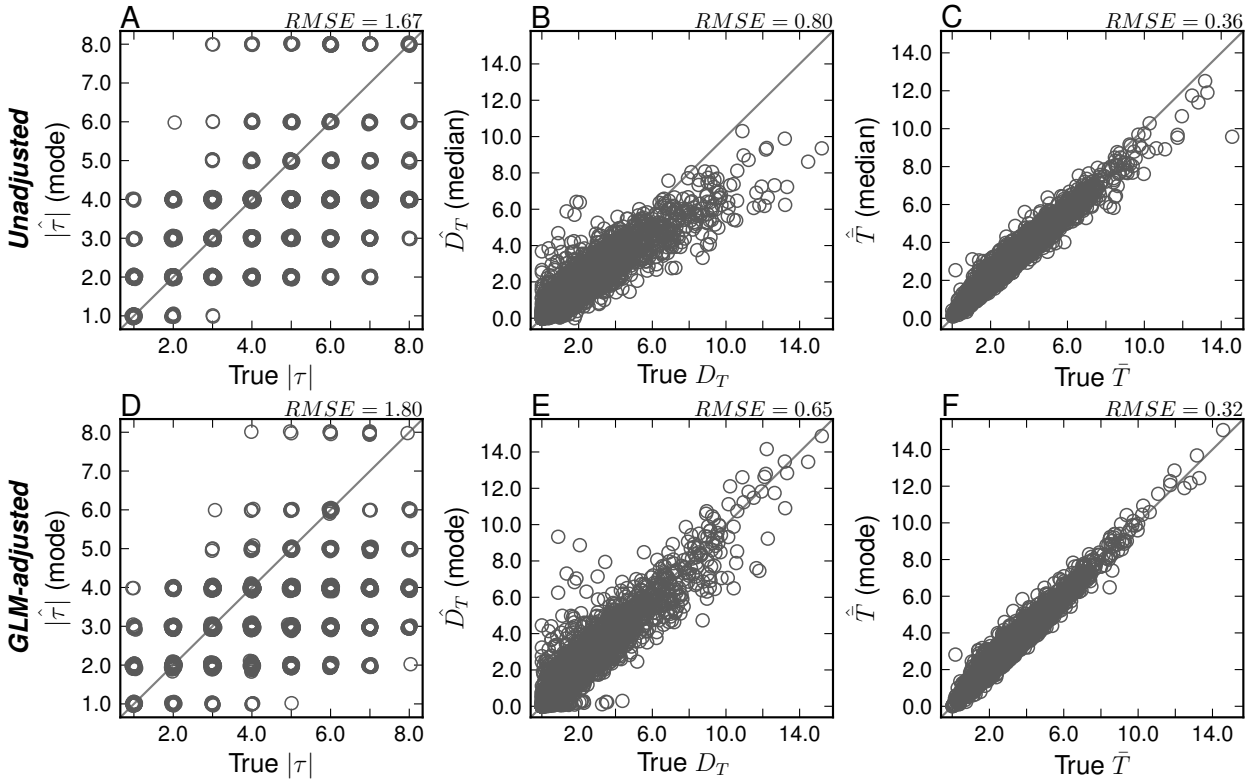

Figure 3.2. Estimation accuracy for model $M_{Uniform}$ when analyzing data generated under M_{DPP}. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

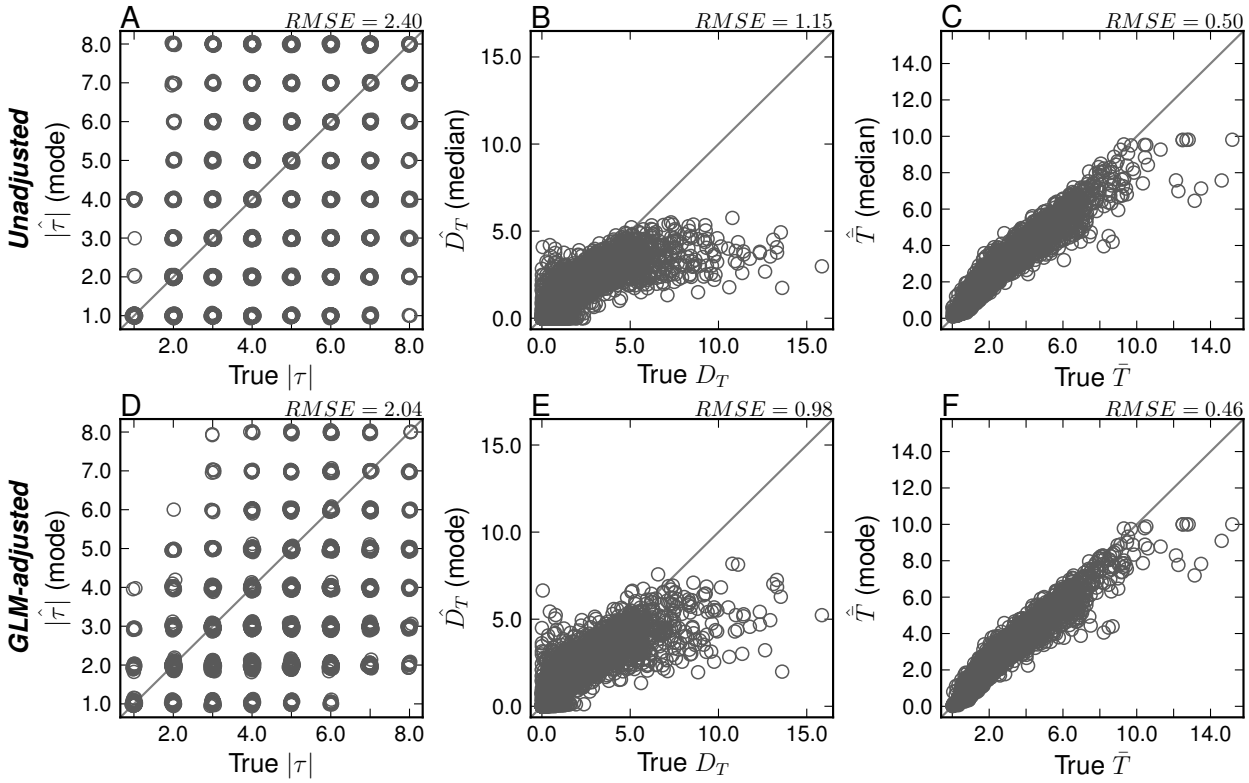

Figure 3.3. Estimation accuracy for model $M_{msBayes}$ when analyzing data generated under M_{DPP}. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

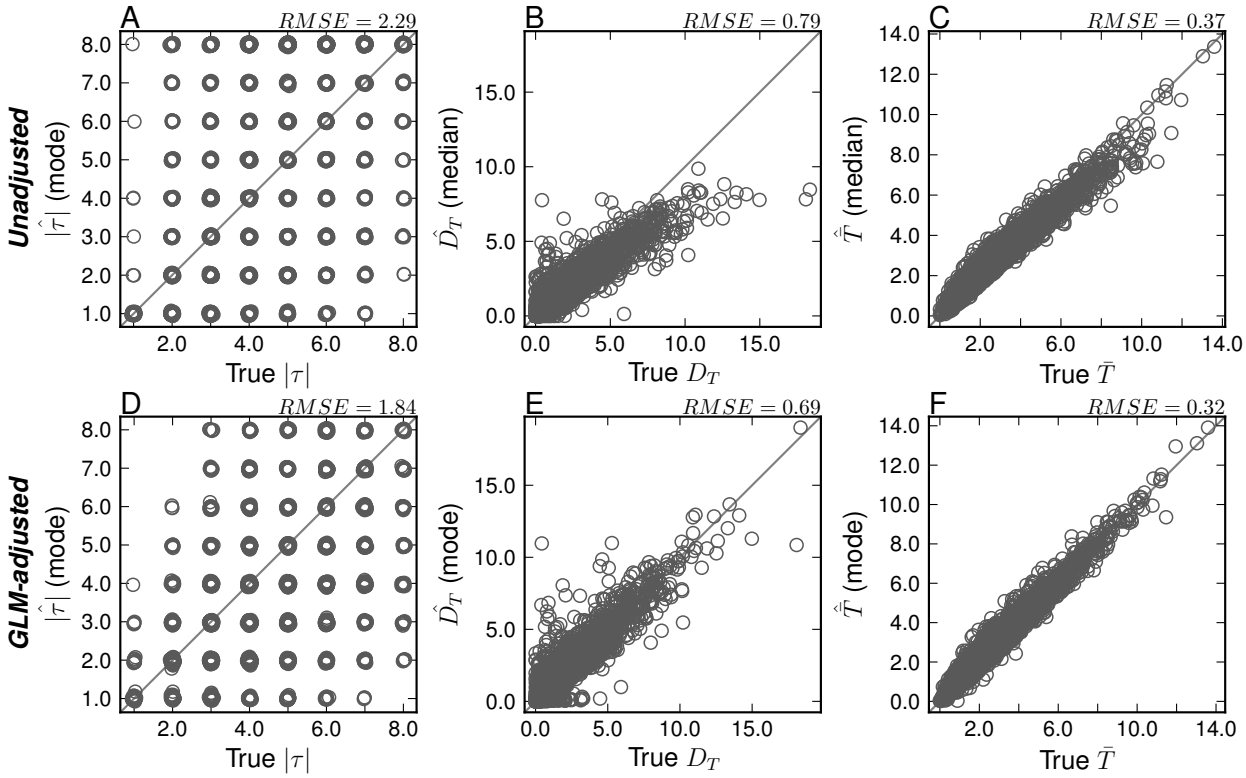

Figure 3.4. Estimation accuracy for model $M_{Ushaped}$ when analyzing data generated under M_{DPP}. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

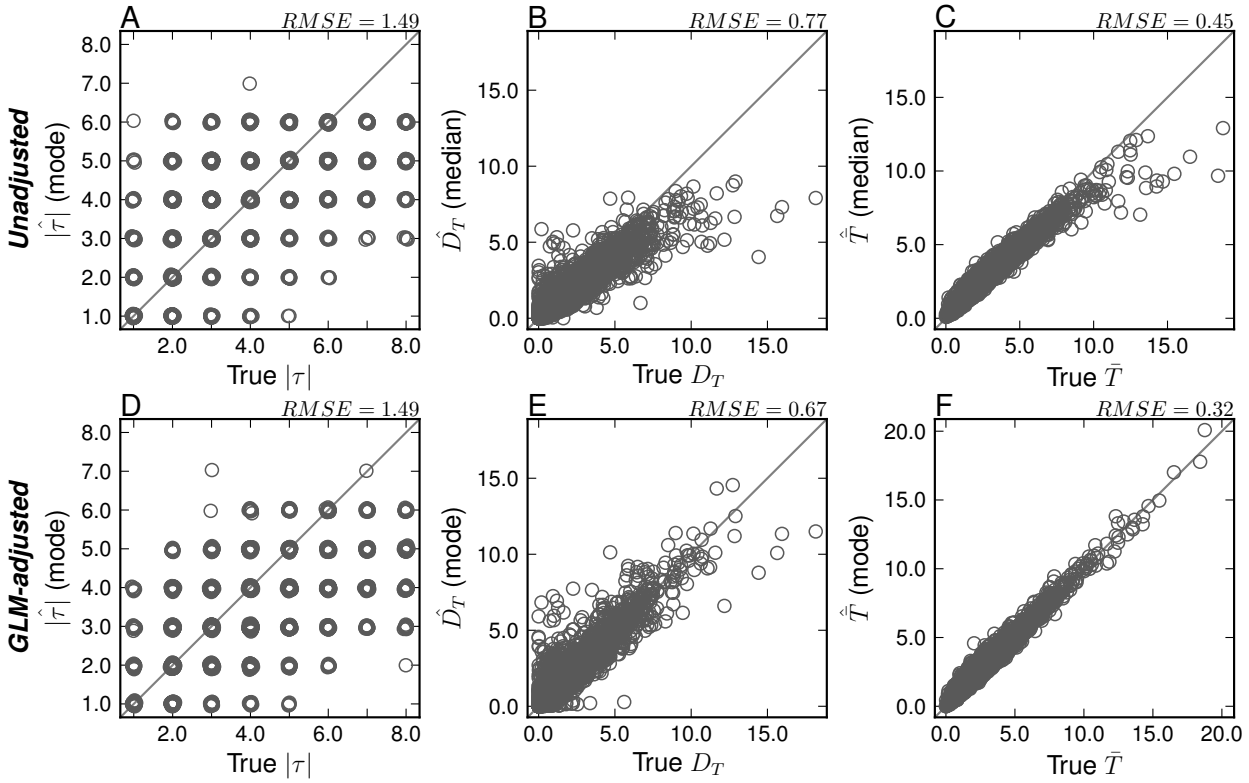

Figure 3.5. Estimation accuracy for model M_{DPP} when analyzing data generated under $M_{Uniform}$. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

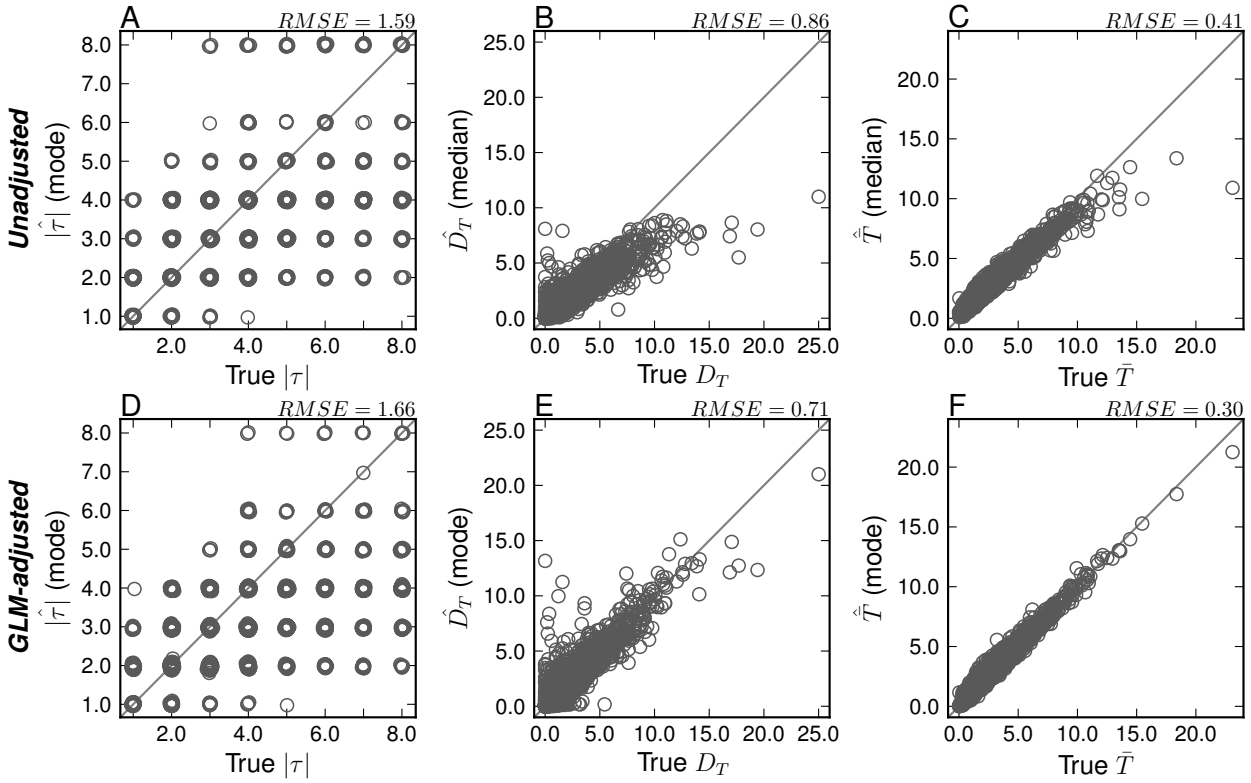

Figure 3.6. Estimation accuracy for model $M_{Uniform}$ when analyzing data generated under $M_{Uniform}$. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

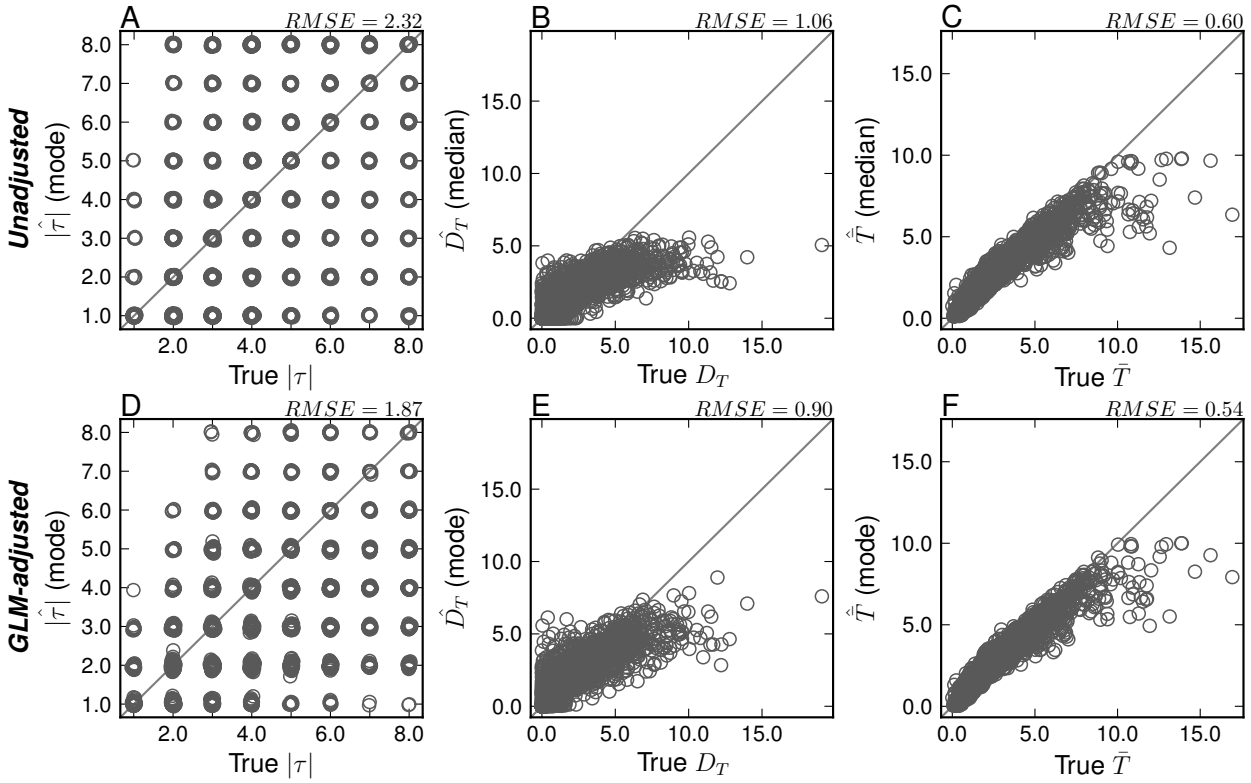

Figure 3.7. Estimation accuracy for model $M_{msBayes}$ when analyzing data generated under $M_{Uniform}$. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

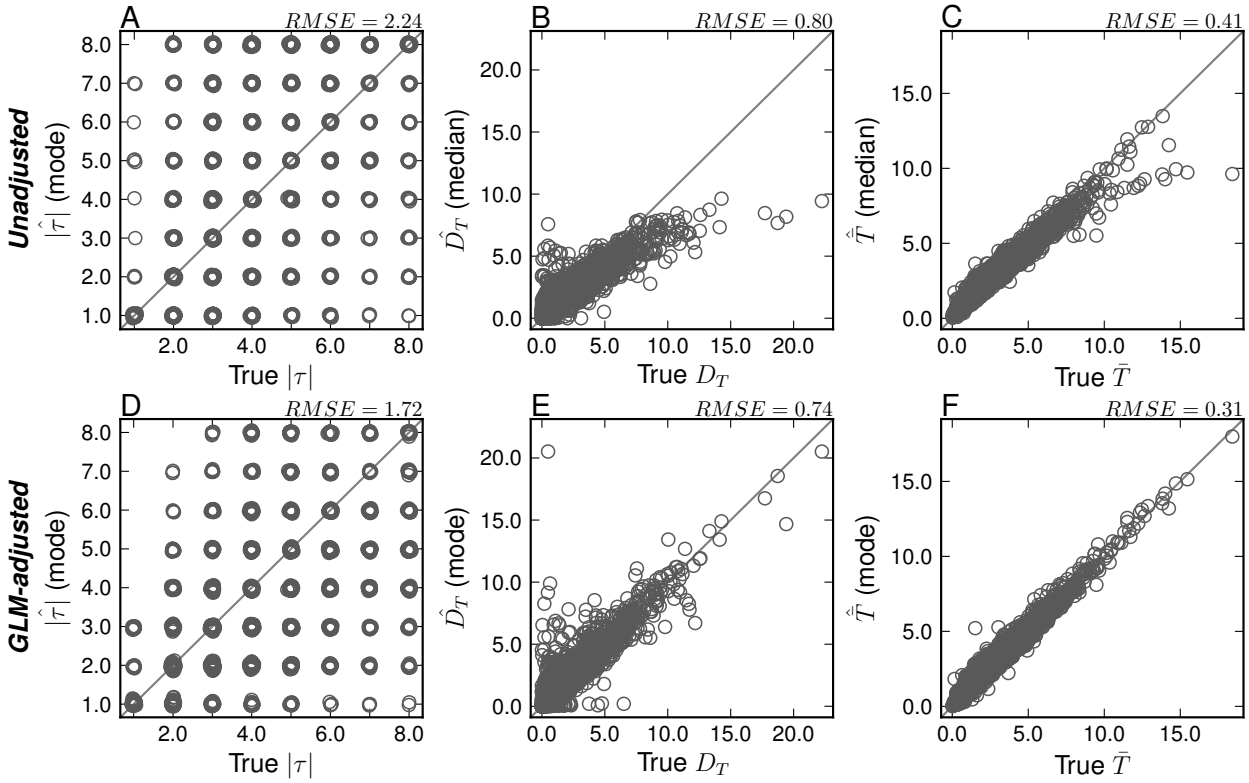

Figure 3.8. Estimation accuracy for model $M_{Ushaped}$ when analyzing data generated under $M_{Uniform}$. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

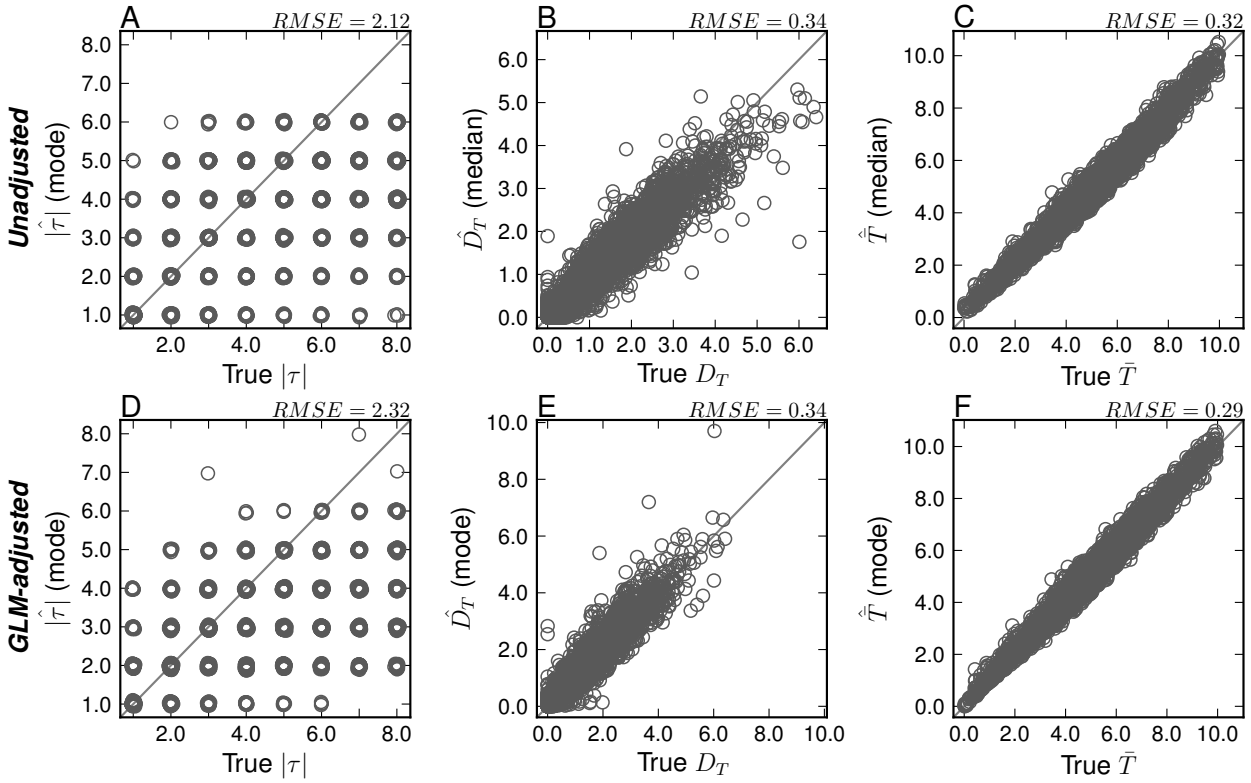

Figure 3.9. Estimation accuracy for model M_{DPP} when analyzing data generated under $M_{msBayes}$. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

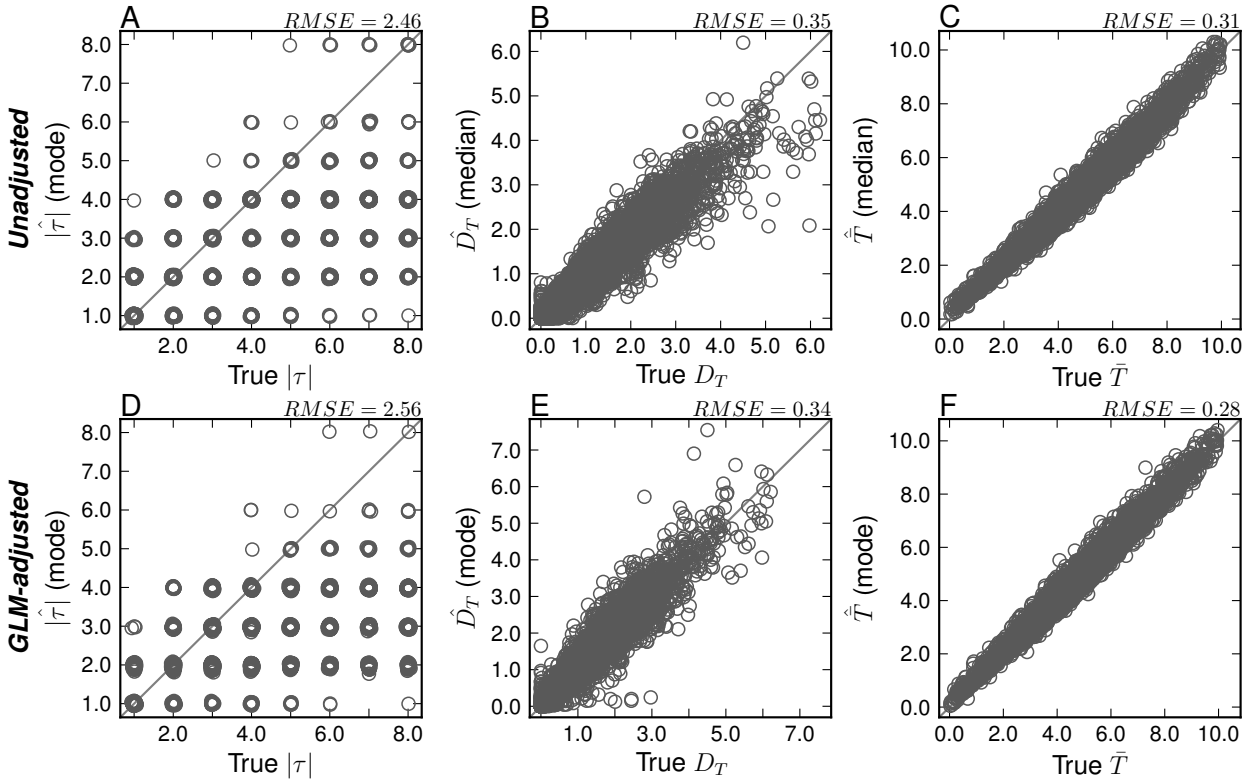

Figure 3.10. Estimation accuracy for model $M_{Uniform}$ when analyzing data generated under $M_{msBayes}$. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

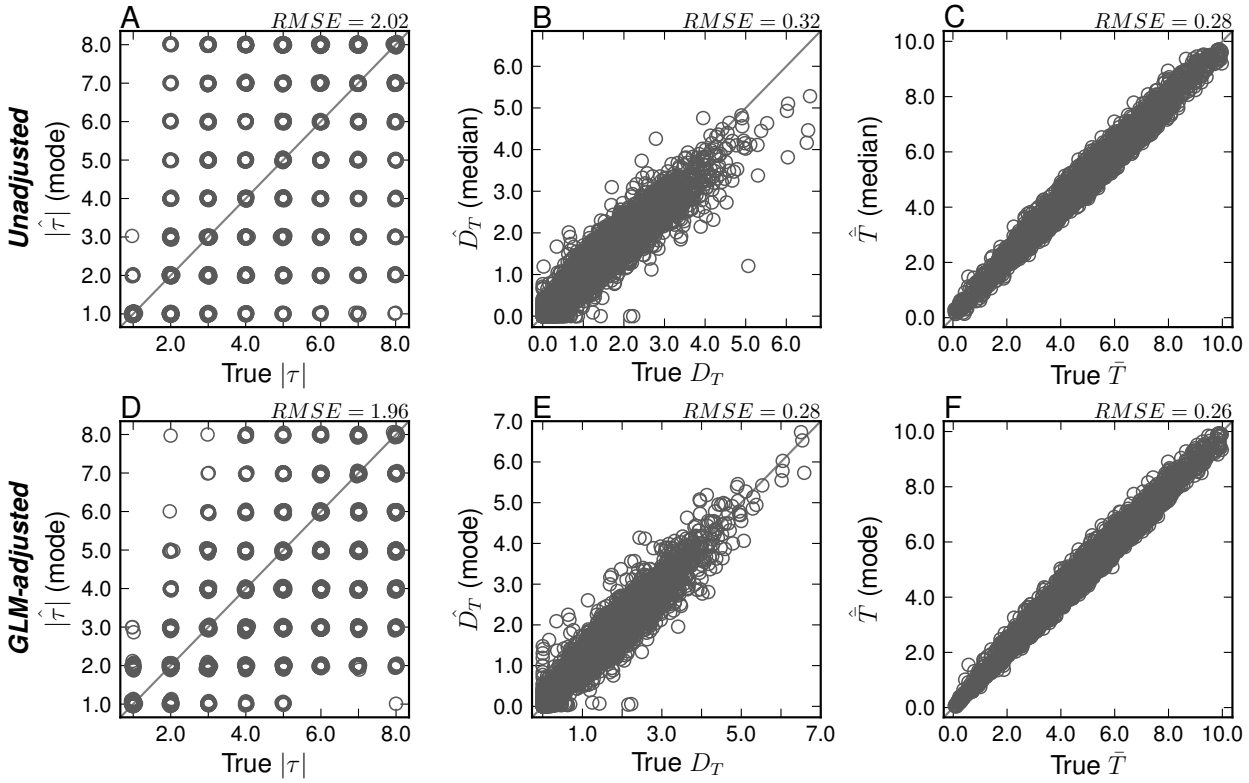

Figure 3.11. Estimation accuracy for model $M_{msBayes}$ when analyzing data generated under $M_{msBayes}$. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

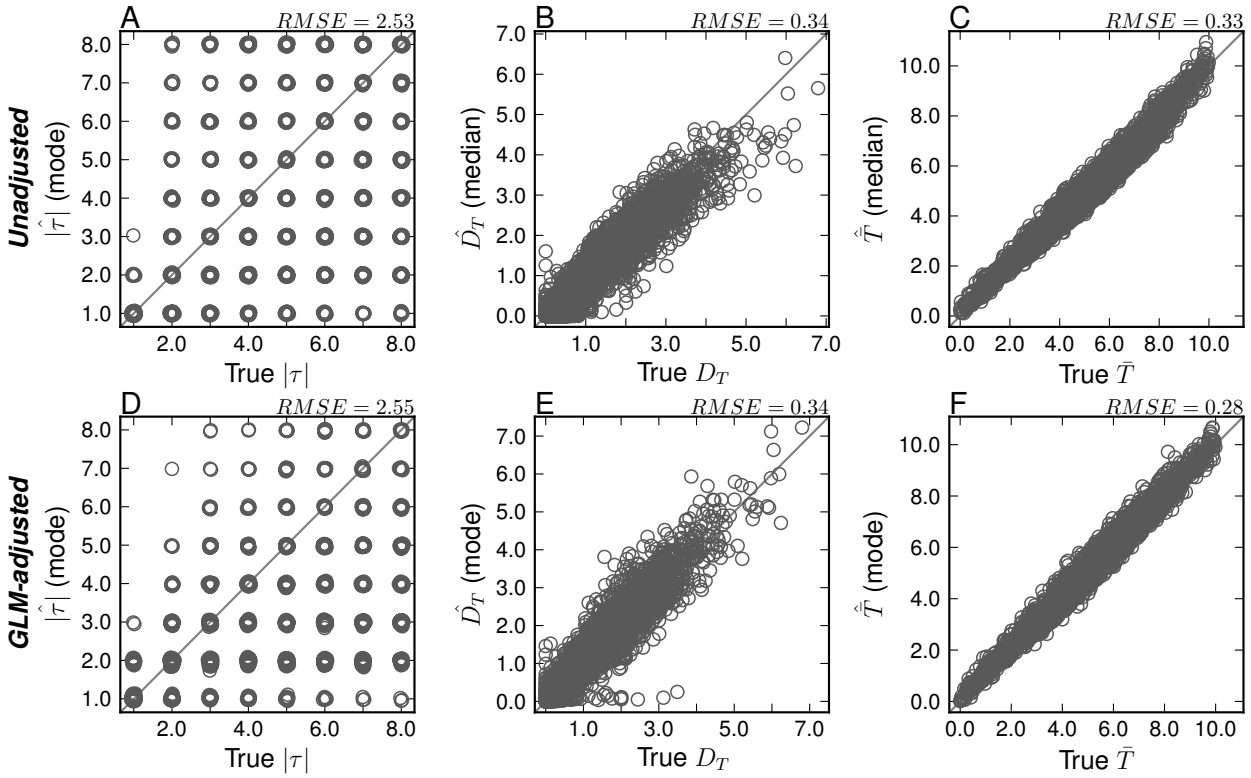

Figure 3.12. Estimation accuracy for model $M_{Ushaped}$ when analyzing data generated under $M_{msBayes}$. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

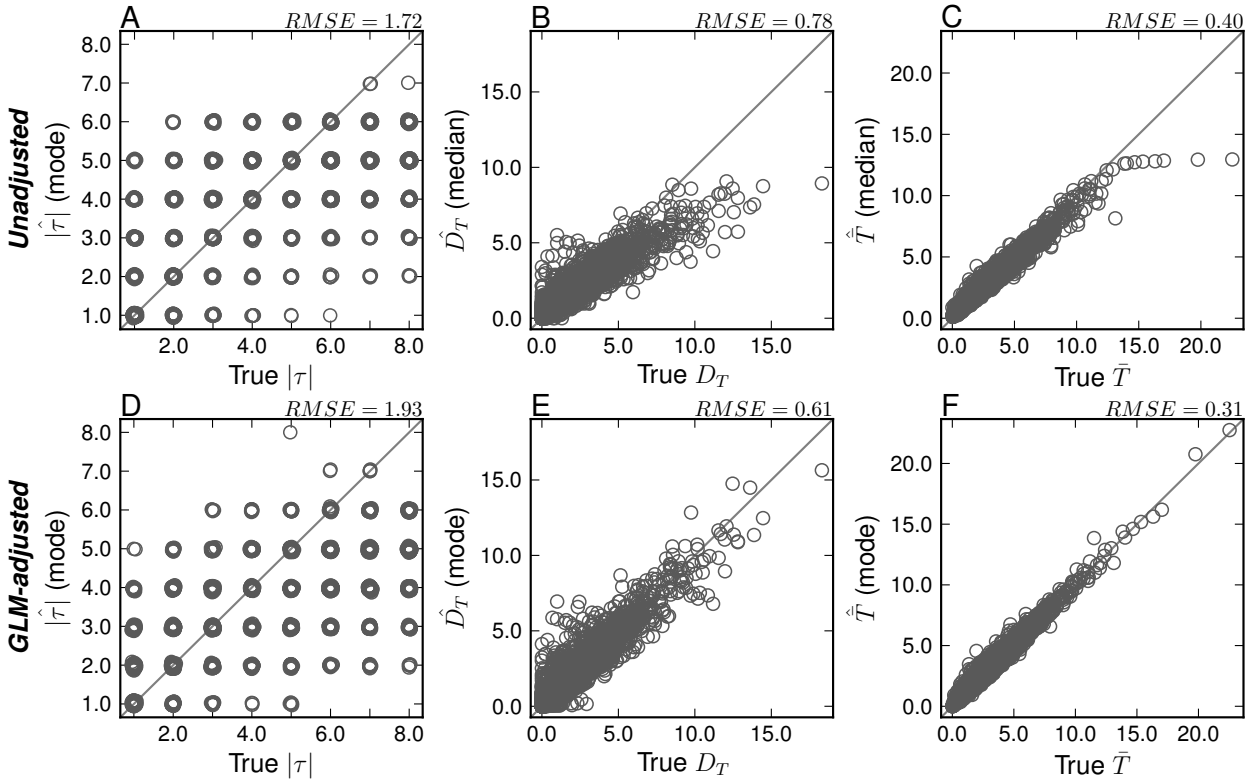

Figure 3.13. Estimation accuracy for model M_{DPP} when analyzing data generated under $M_{Ushaped}$. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

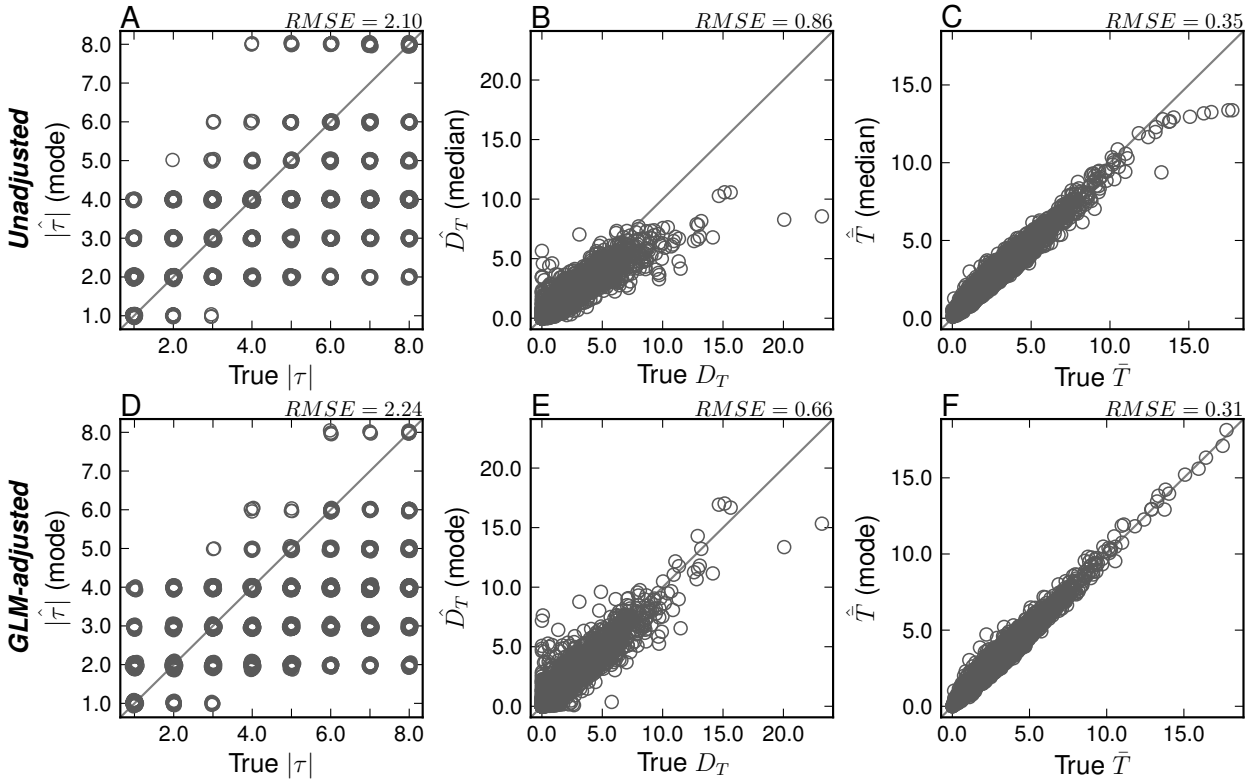

Figure 3.14. Estimation accuracy for model $M_{Uniform}$ when analyzing data generated under $M_{Ushaped}$. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

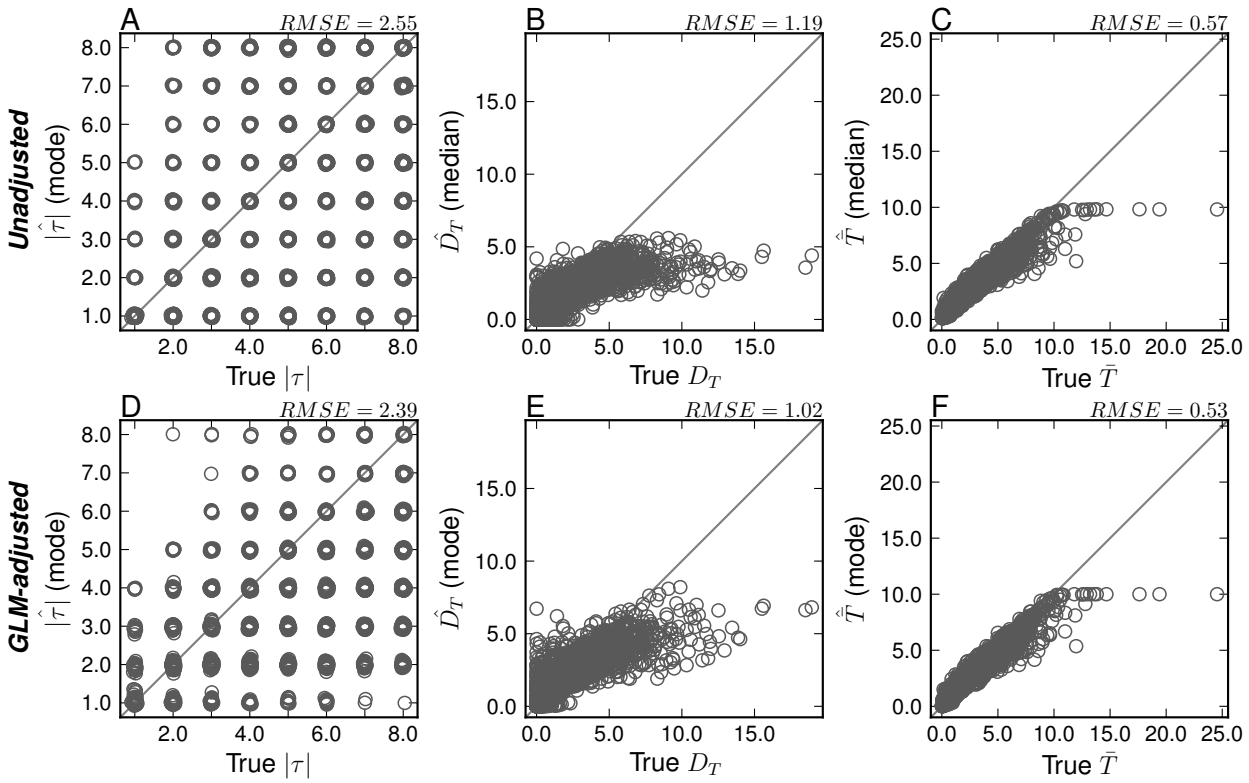

Figure 3.15. Estimation accuracy for model $M_{msBayes}$ when analyzing data generated under $M_{Ushaped}$. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

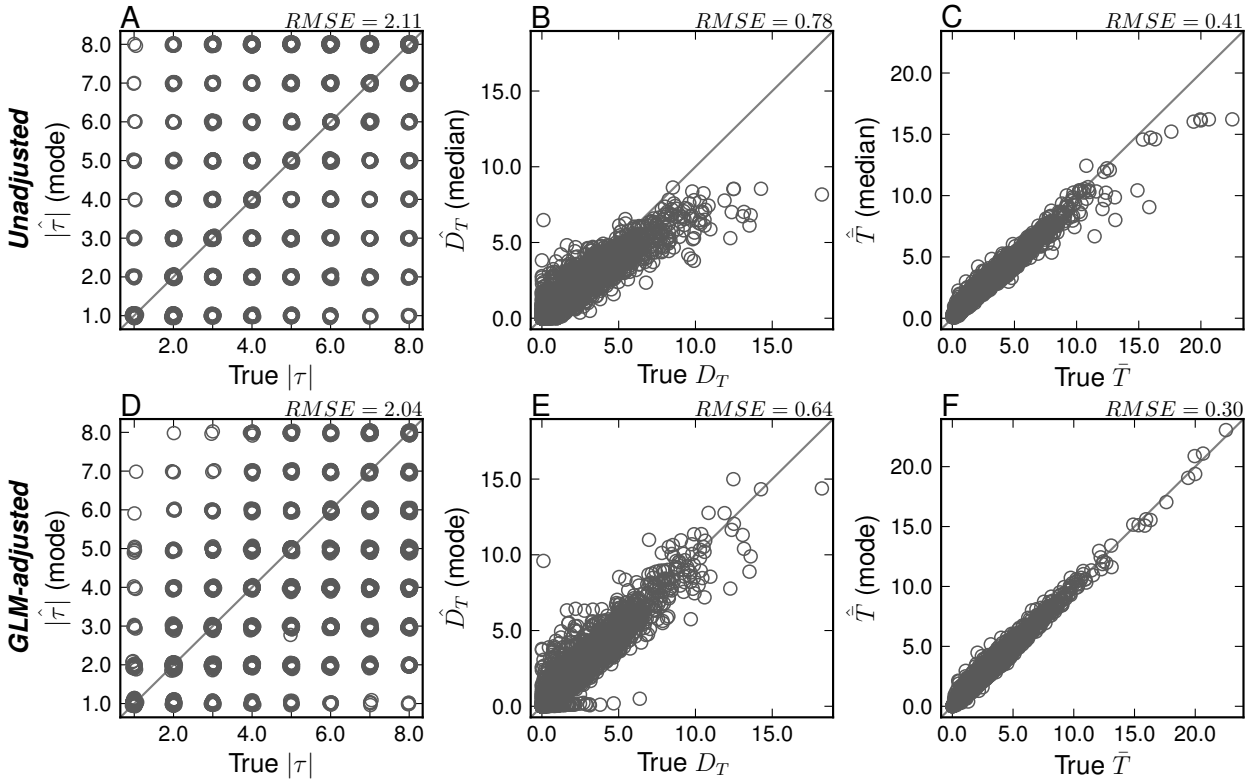

Figure 3.16. Estimation accuracy for model $M_{Ushaped}$ when analyzing data generated under $M_{Ushaped}$. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

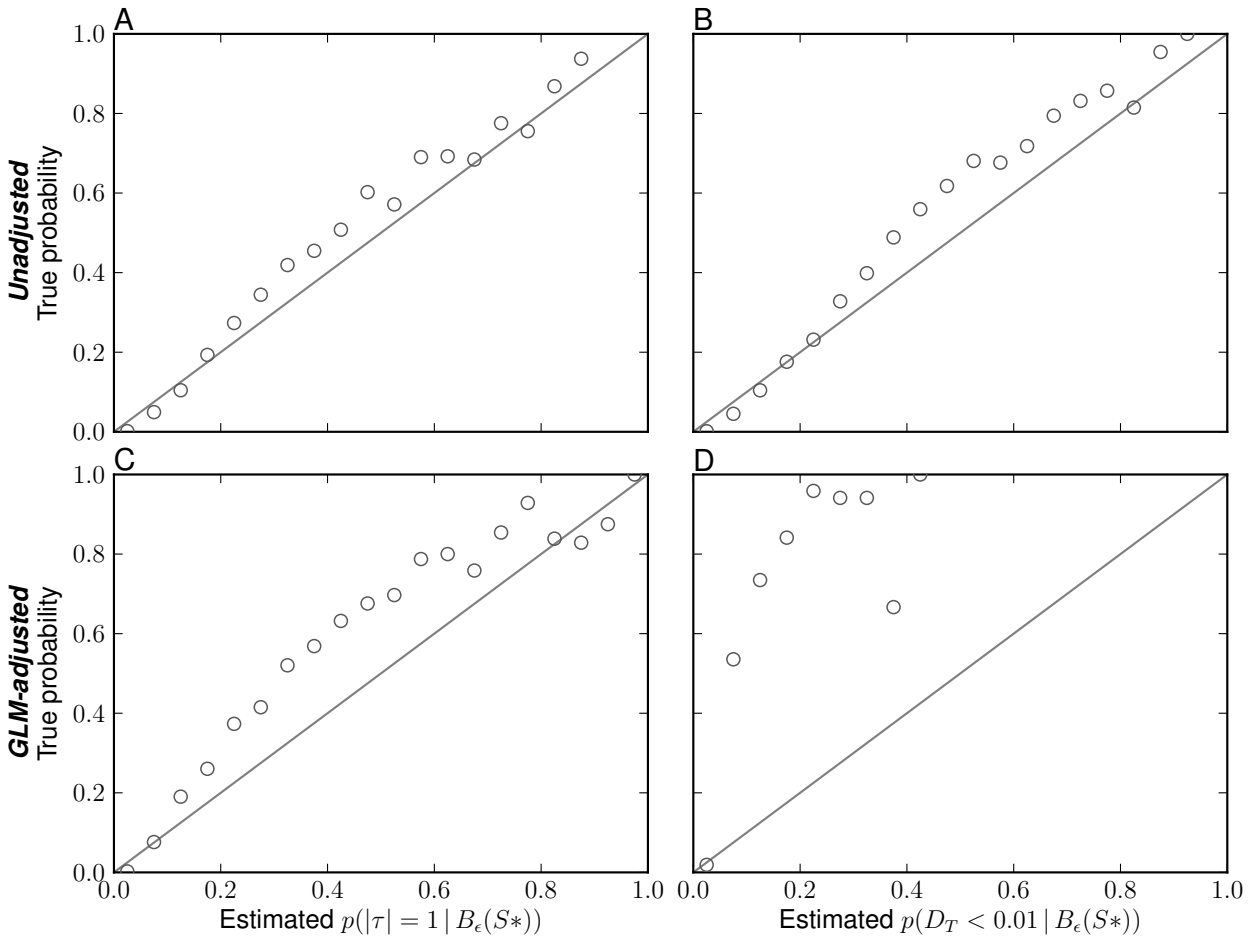

Figure 3.17. Model-choice accuracy for model M_{DPP} when analyzing data generated under M_{DPP}. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

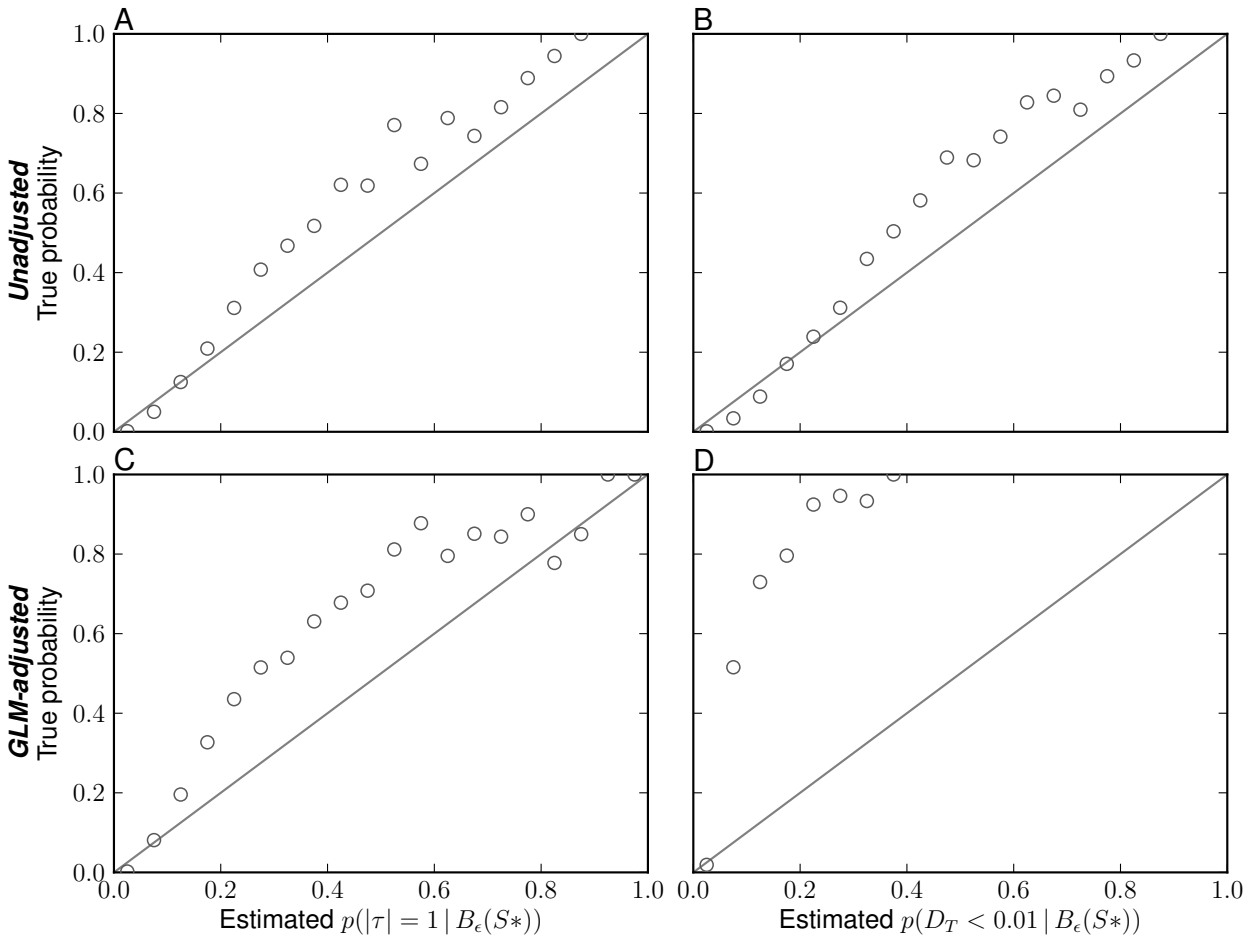

Figure 3.18. Model-choice accuracy for model $M_{Uniform}$ when analyzing data generated under M_{DPP}. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

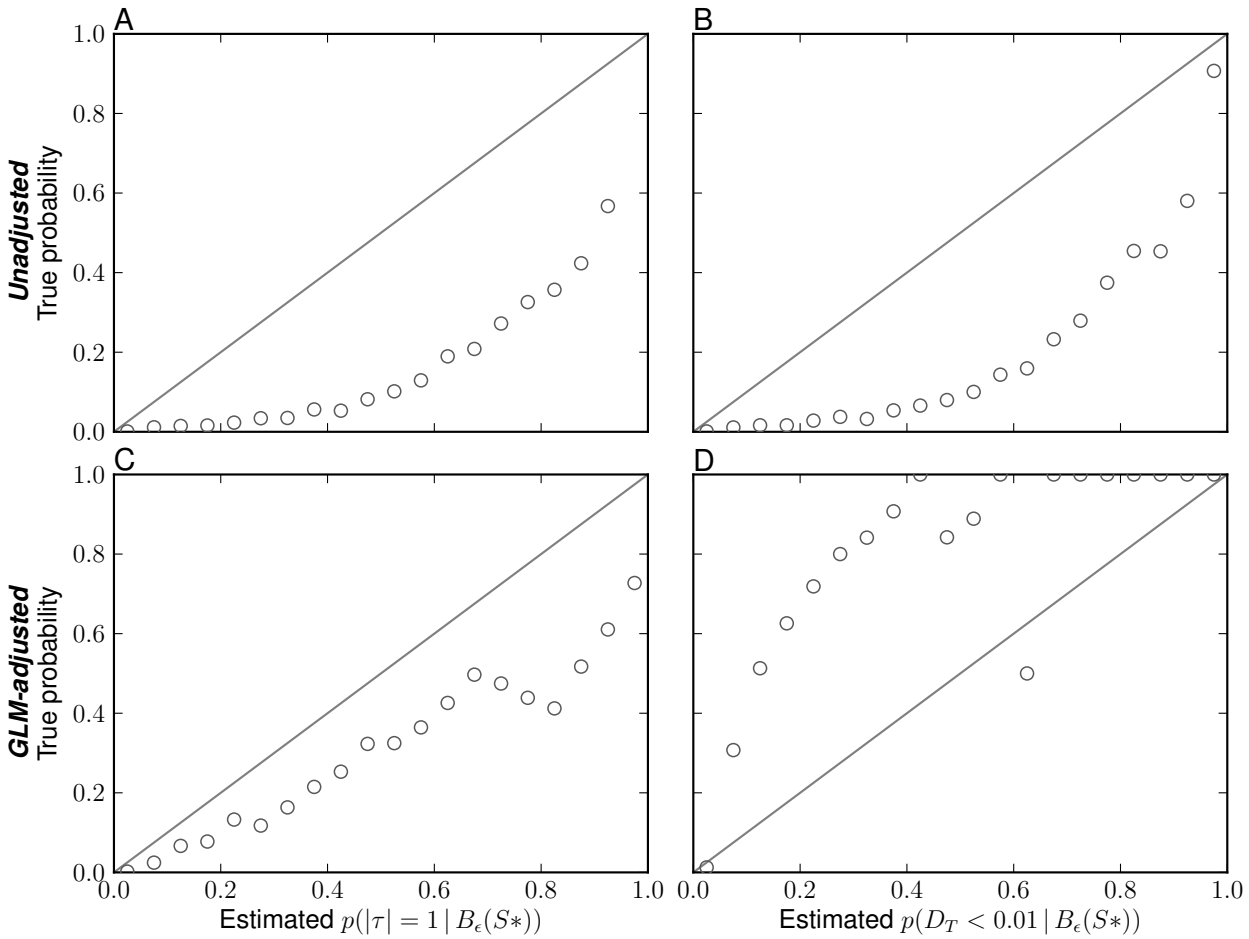

Figure 3.19. Model-choice accuracy for model $M_{msBayes}$ when analyzing data generated under M_{DPP}. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

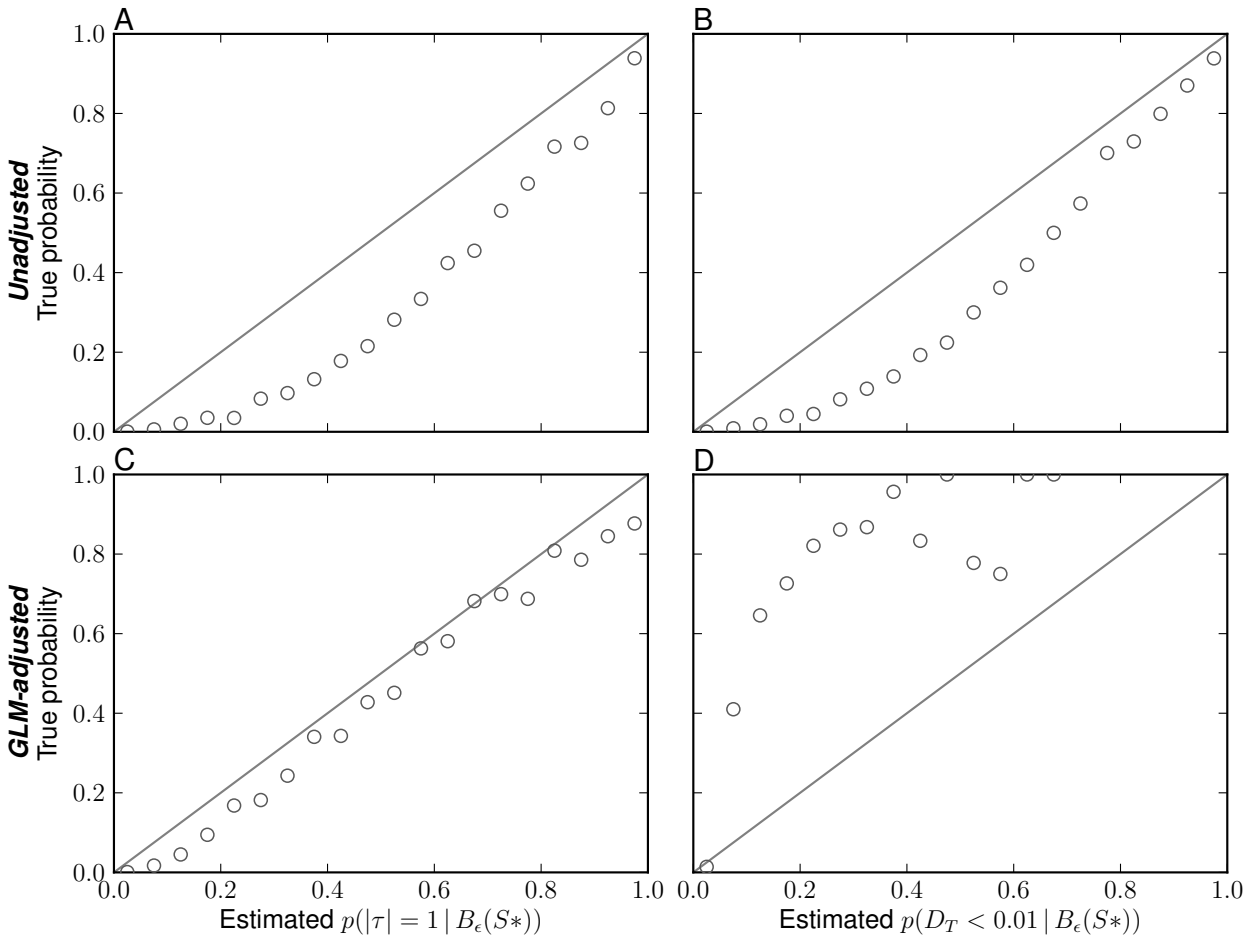

Figure 3.20. Model-choice accuracy for model $M_{Ushaped}$ when analyzing data generated under M_{DPP}. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

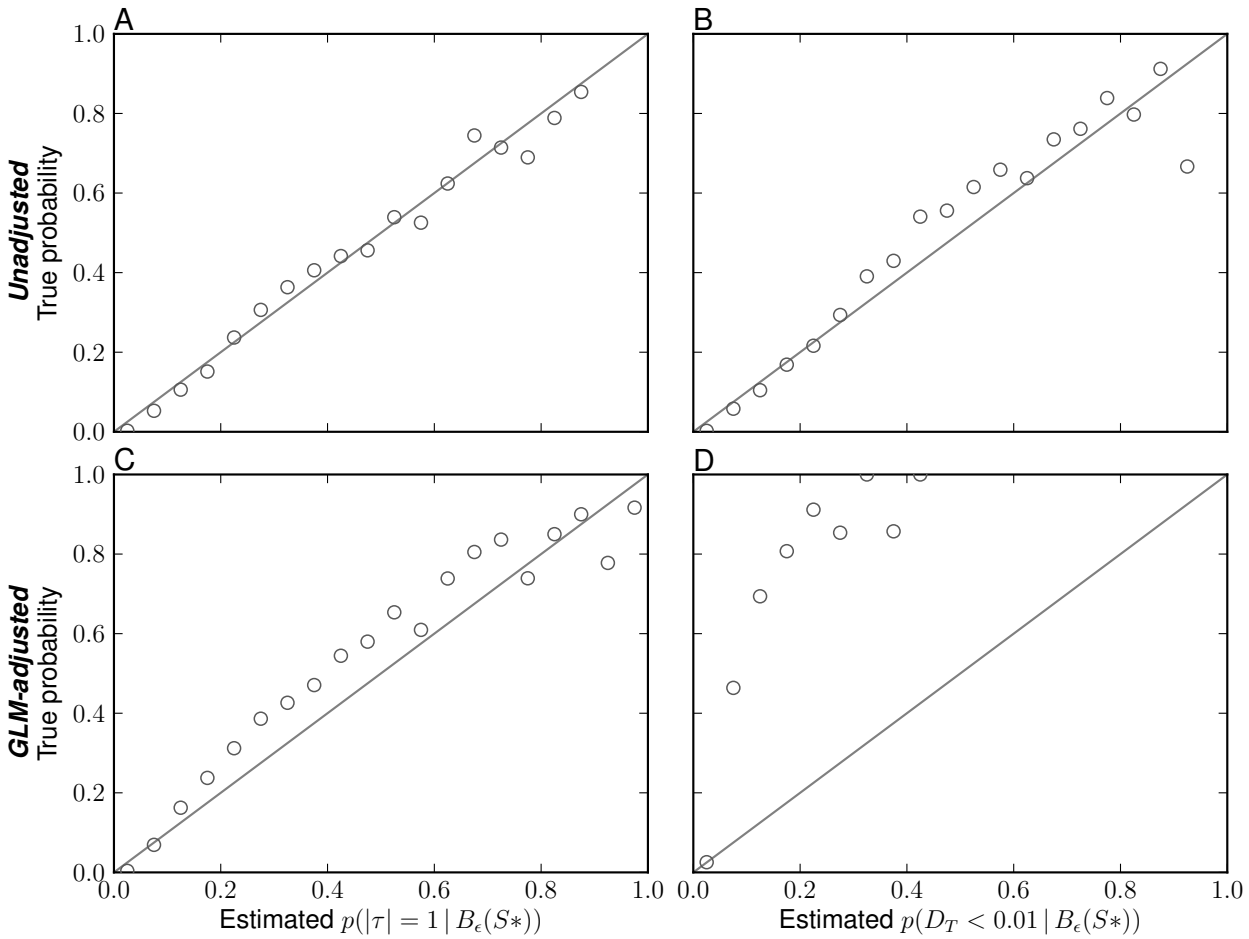

Figure 3.21. Model-choice accuracy for model M_{DPP} when analyzing data generated under $M_{Uniform}$. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

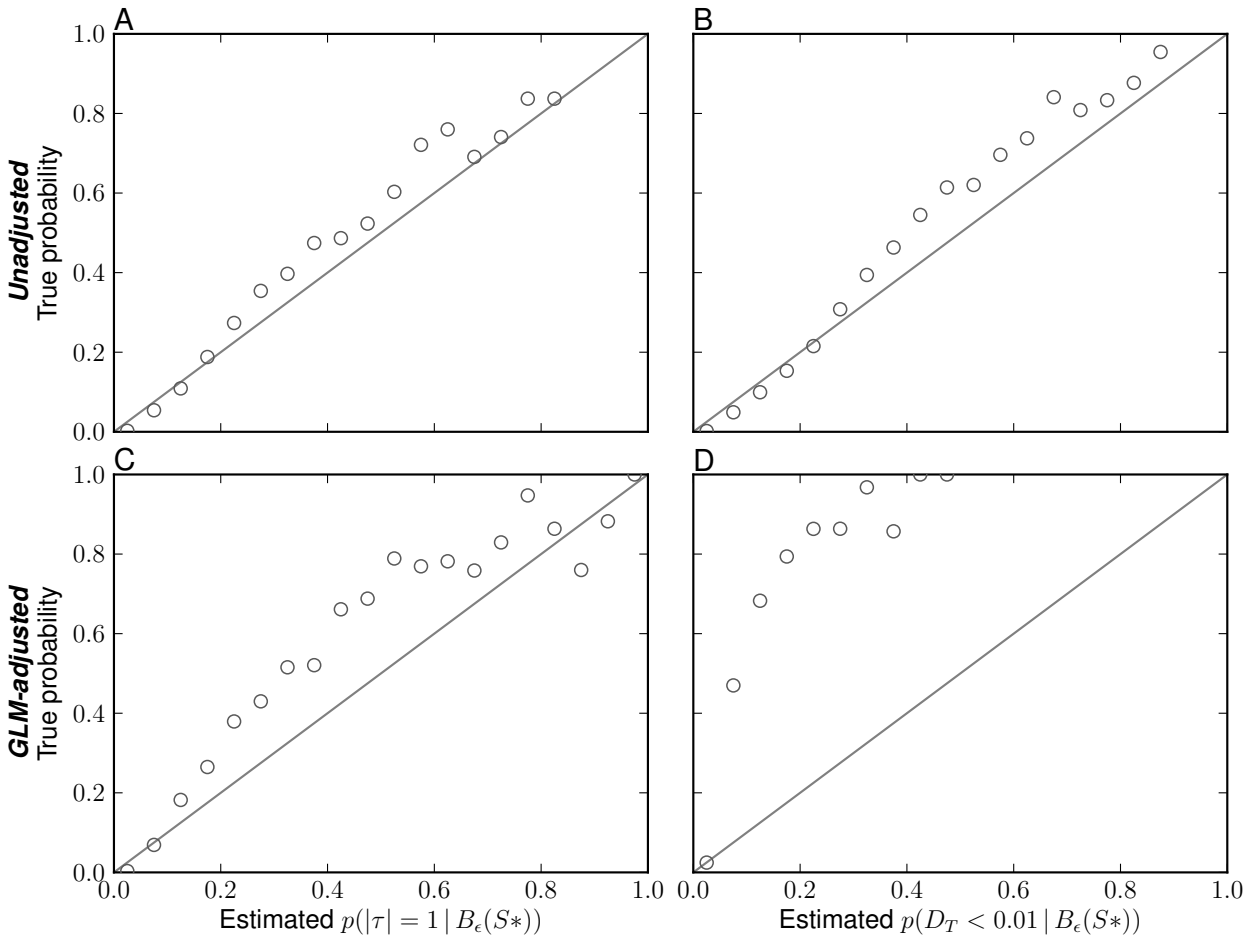

Figure 3.22. Model-choice accuracy for model $M_{Uniform}$ when analyzing data generated under $M_{Uniform}$. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

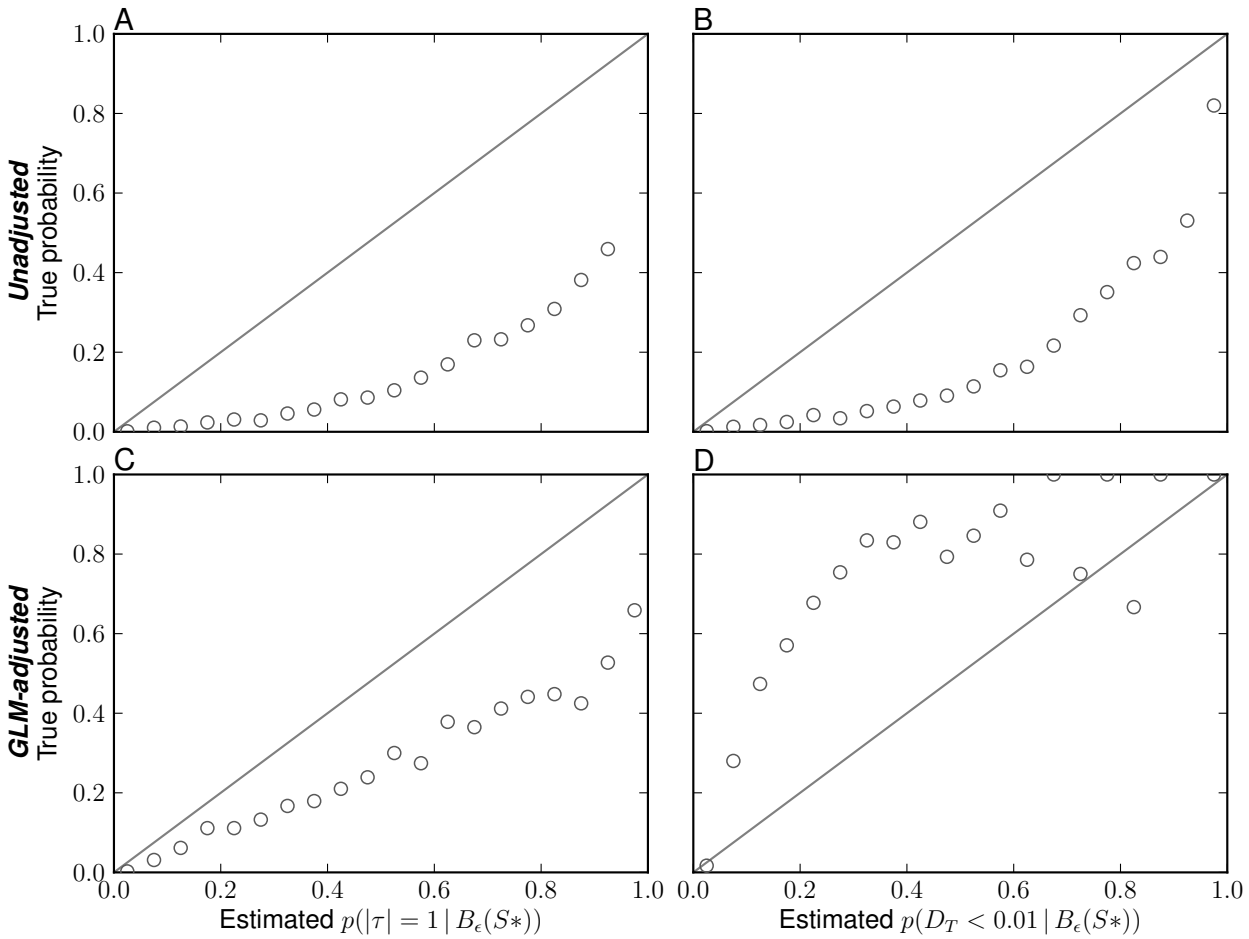

Figure 3.23. Model-choice accuracy for model $M_{msBayes}$ when analyzing data generated under $M_{Uniform}$. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

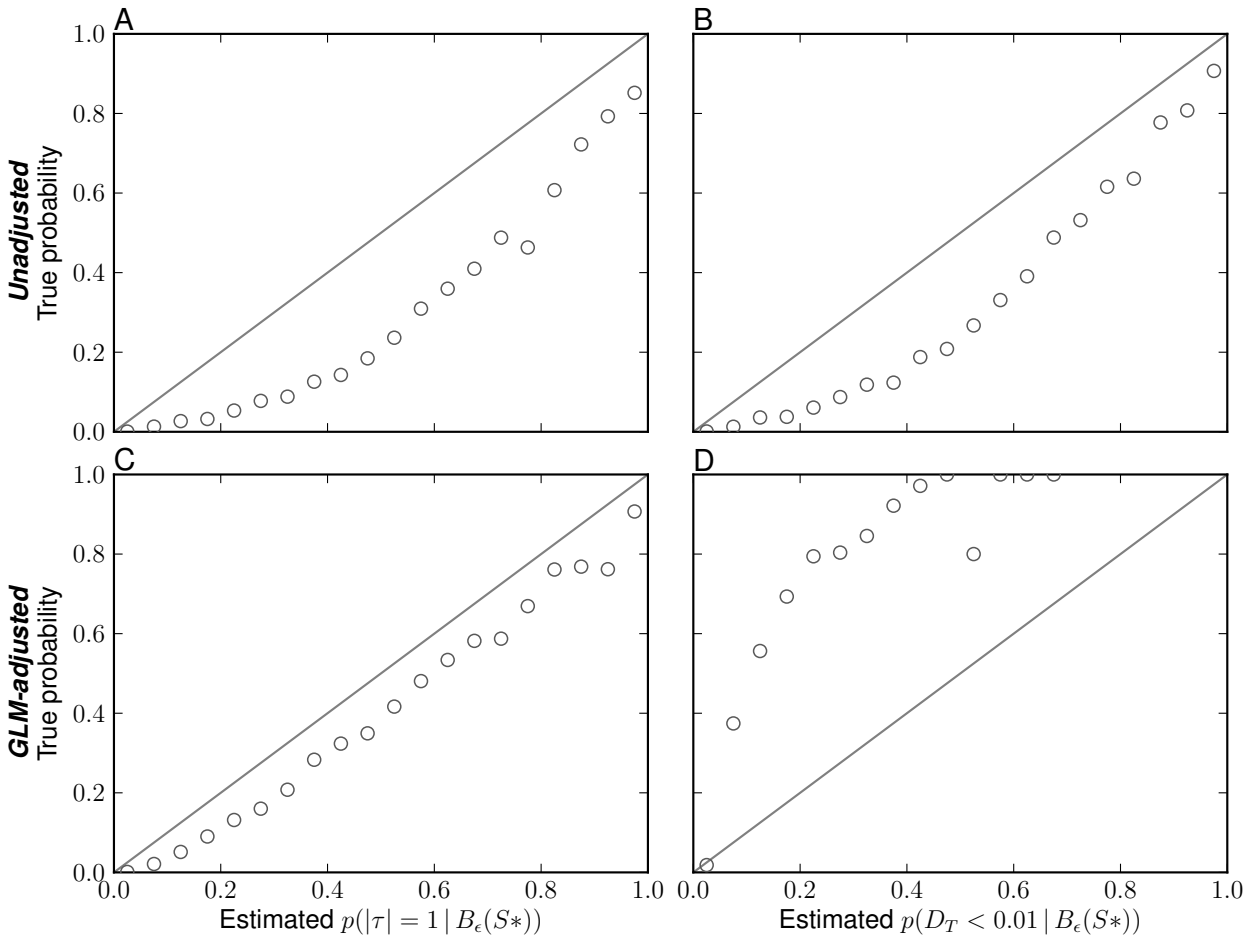

Figure 3.24. Model-choice accuracy for model $M_{Ushaped}$ when analyzing data generated under $M_{Uniform}$. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

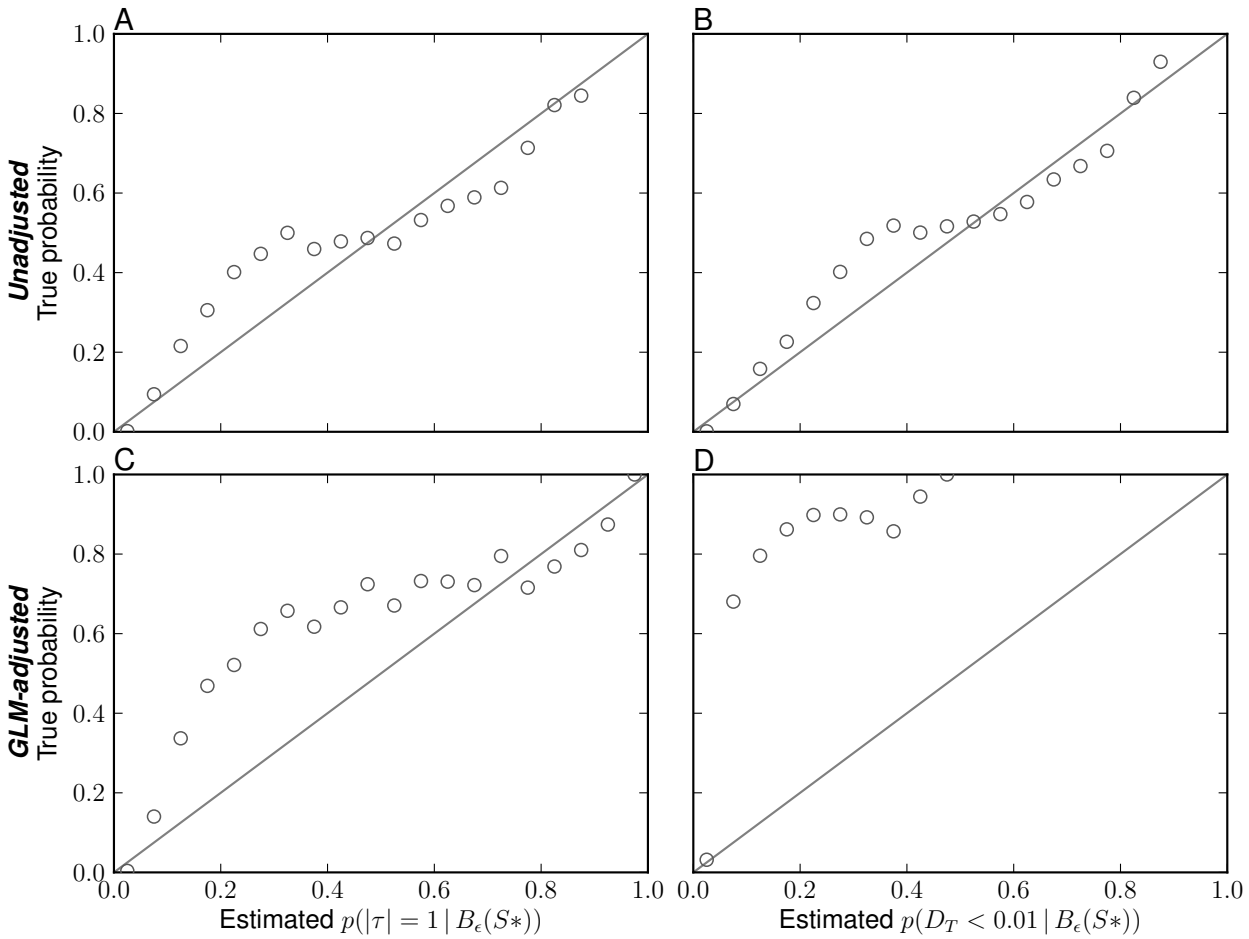

Figure 3.25. Model-choice accuracy for model M_{DPP} when analyzing data generated under $M_{msBayes}$. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

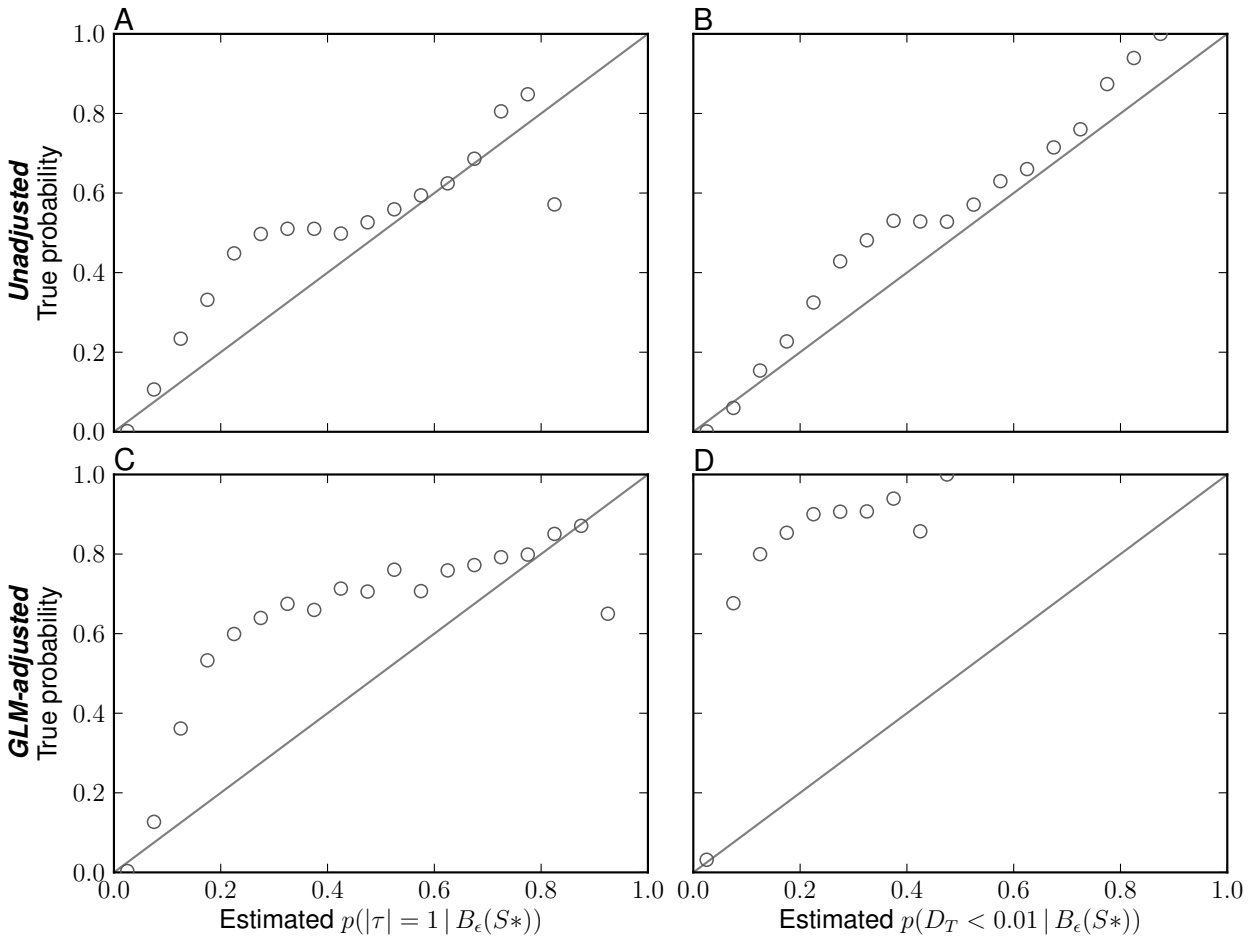

Figure 3.26. Model-choice accuracy for model $M_{Uniform}$ when analyzing data generated under $M_{msBayes}$. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

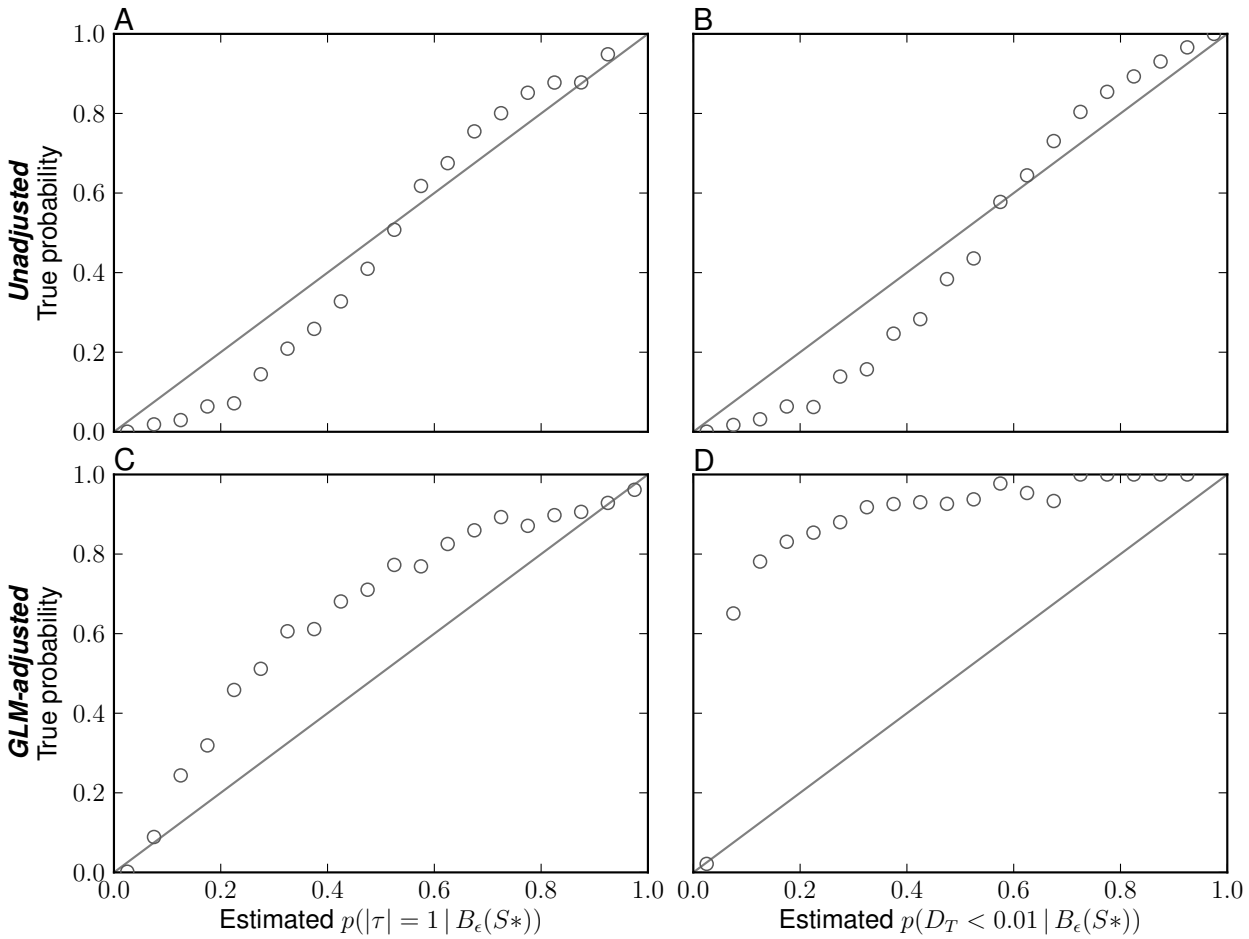

Figure 3.27. Model-choice accuracy for model $M_{msBayes}$ when analyzing data generated under $M_{msBayes}$. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

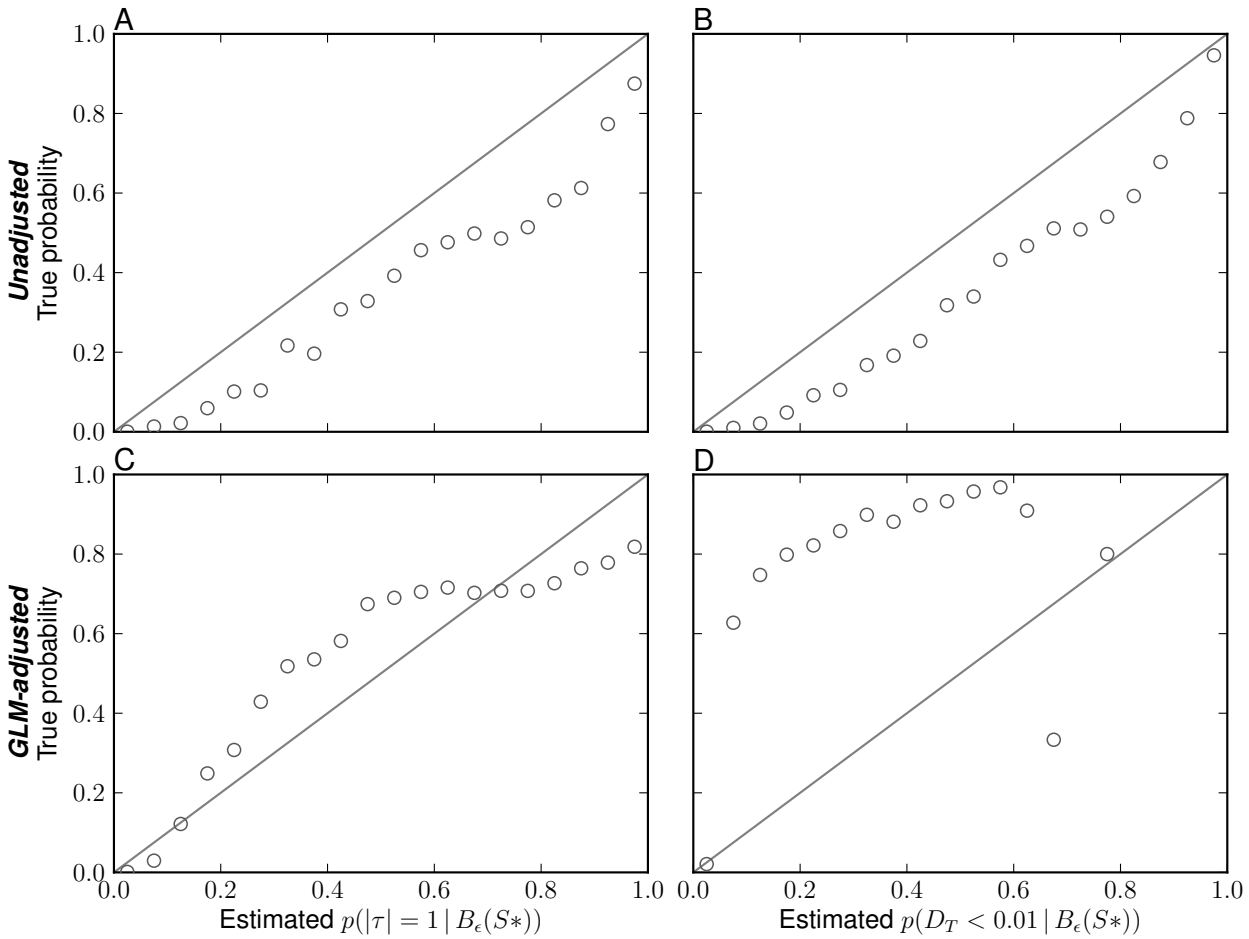

Figure 3.28. Model-choice accuracy for model $M_{Ushaped}$ when analyzing data generated under $M_{msBayes}$. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

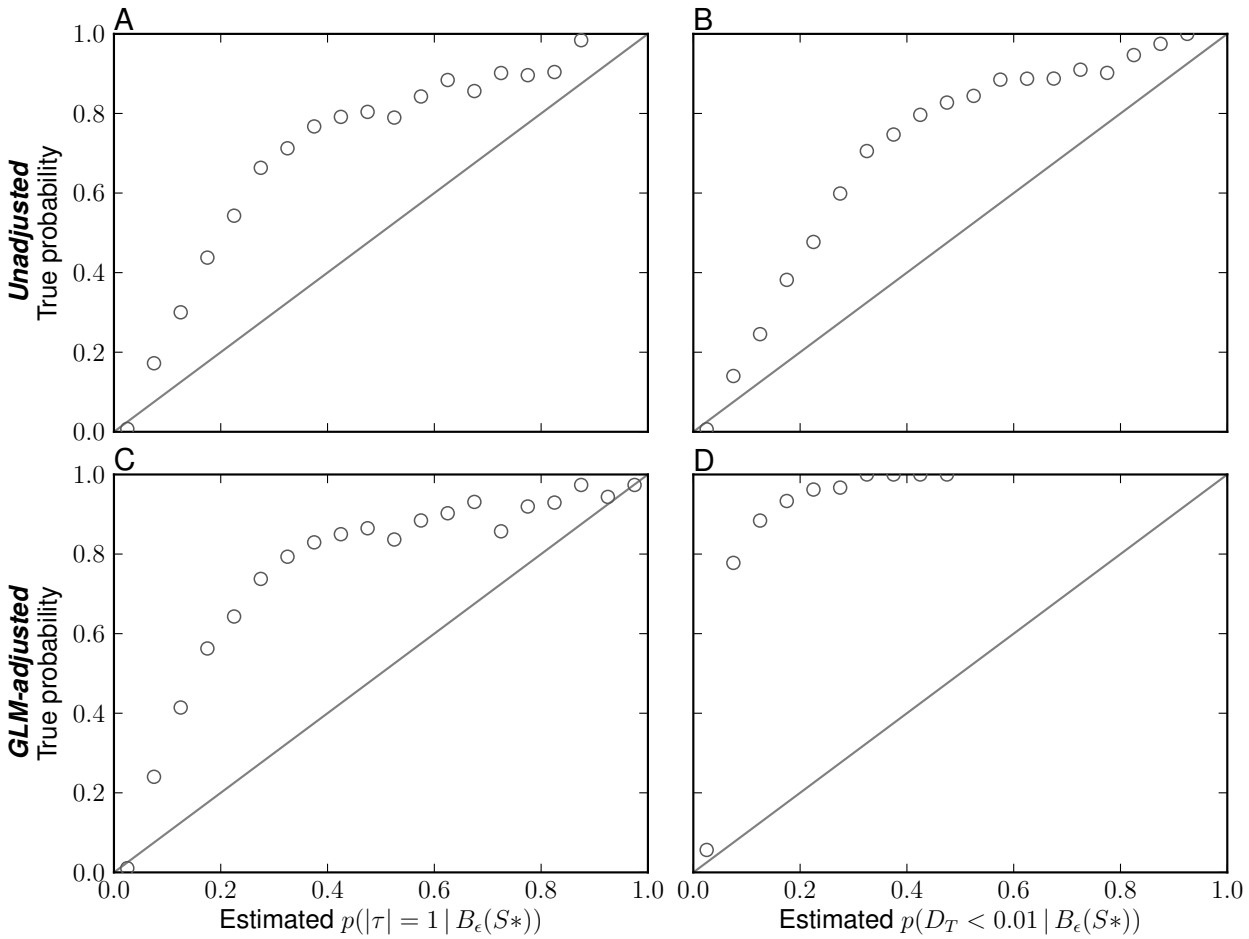

Figure 3.29. Model-choice accuracy for model M_{DPP} when analyzing data generated under $M_{Ushaped}$. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

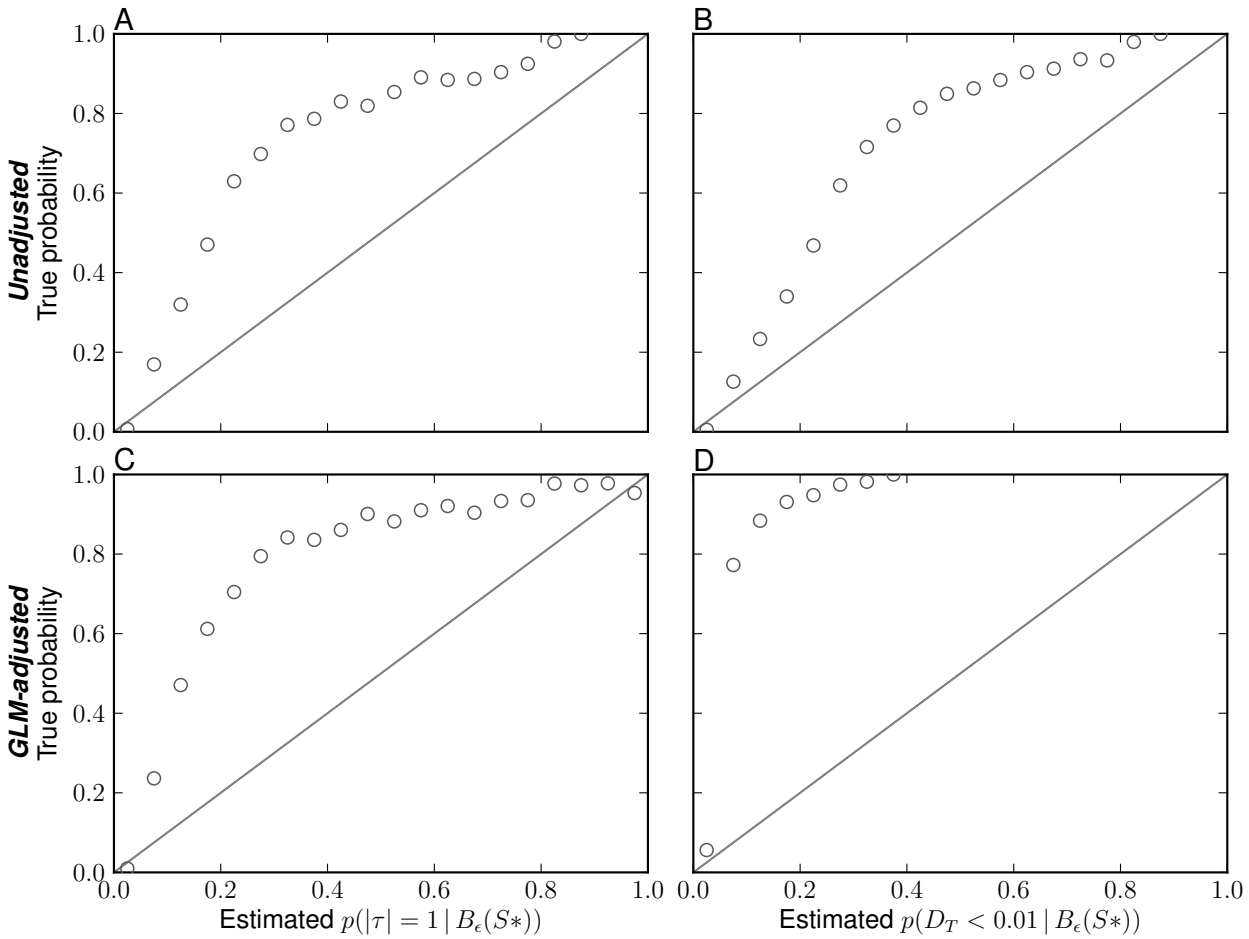

Figure 3.30. Model-choice accuracy for model $M_{Uniform}$ when analyzing data generated under $M_{Ushaped}$. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

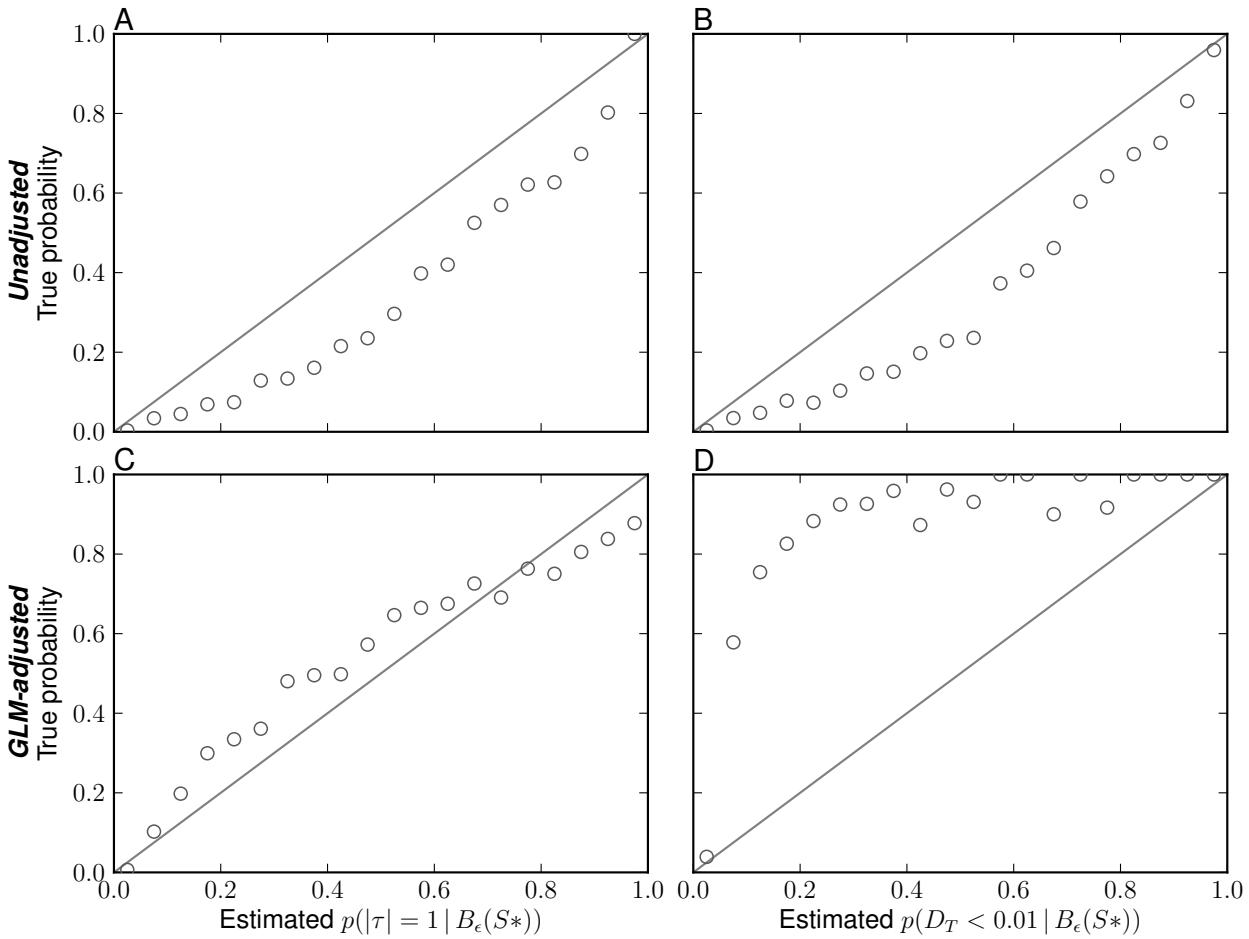

Figure 3.31. Model-choice accuracy for model $M_{msBayes}$ when analyzing data generated under $M_{Ushaped}$. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

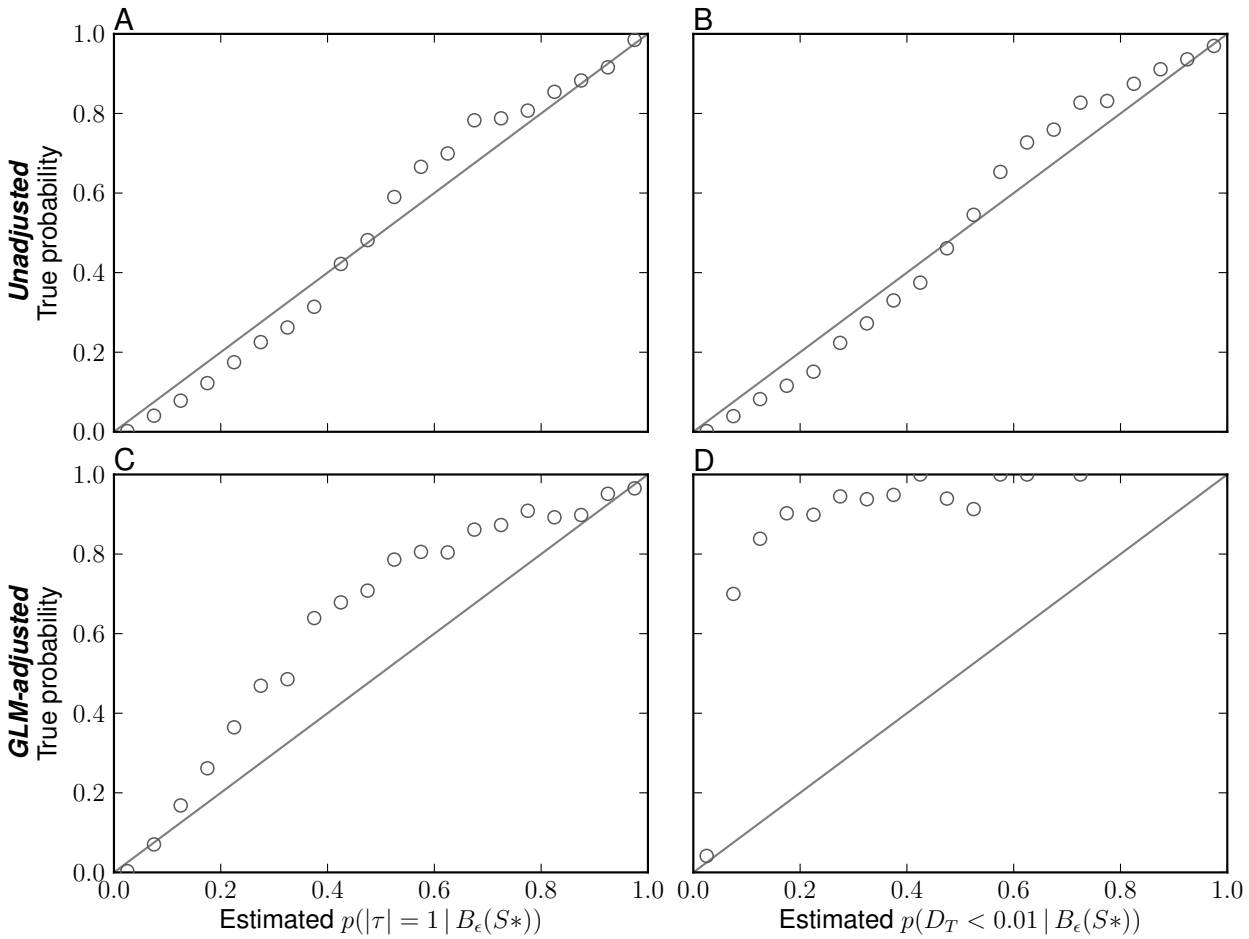

Figure 3.32. Model-choice accuracy for model $M_{Ushaped}$ when analyzing data generated under $M_{Ushaped}$. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

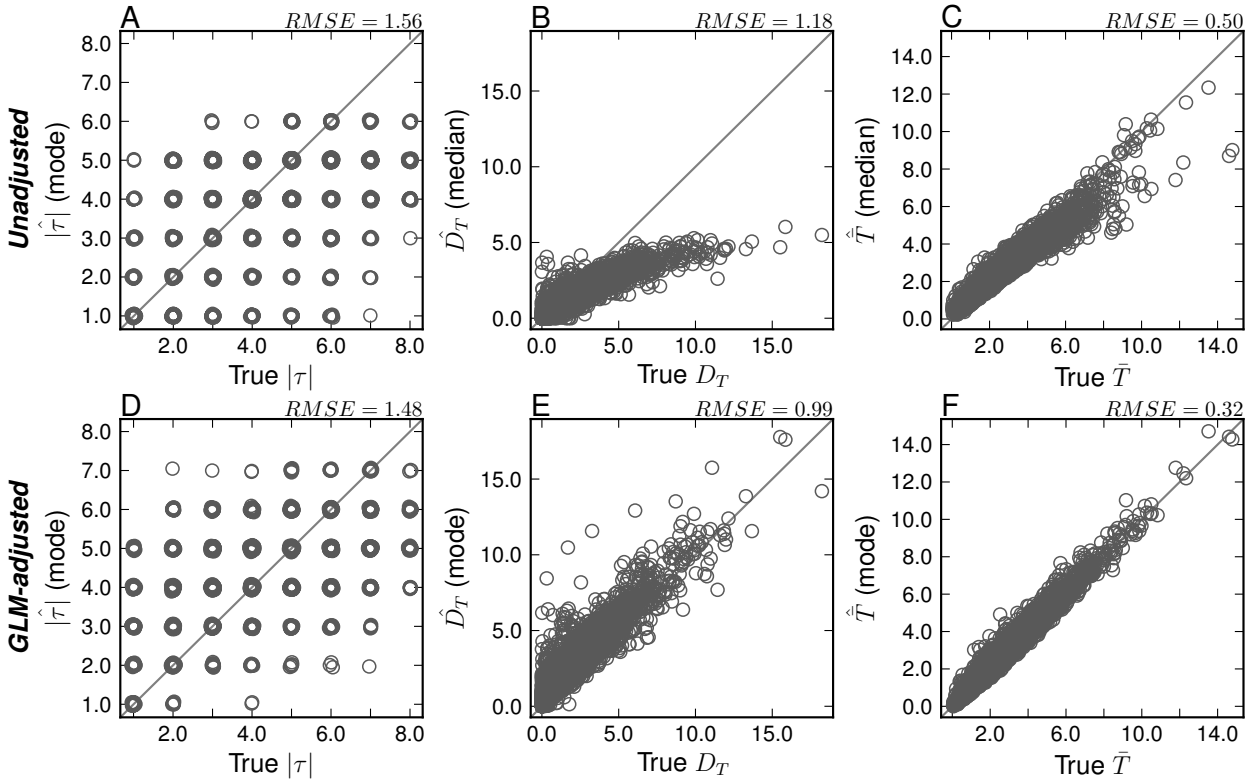

Figure 3.33. Estimation accuracy for model M°_{DPP} when analyzing data generated under M°_{DPP}. A random sample of 5000 posterior estimates (from 50,000) are plotted, including both (A, B, & C) unadjusted and (D, E, & F) GLM-regression-adjusted estimates. Normal random variates ($N(0, 0.005)$) have been added to the estimates and true values of $|\boldsymbol{\tau}|$ (A & D) to reduce overlap of plot symbols. The root mean square error (RMSE) calculated from the 5000 estimates is provided.

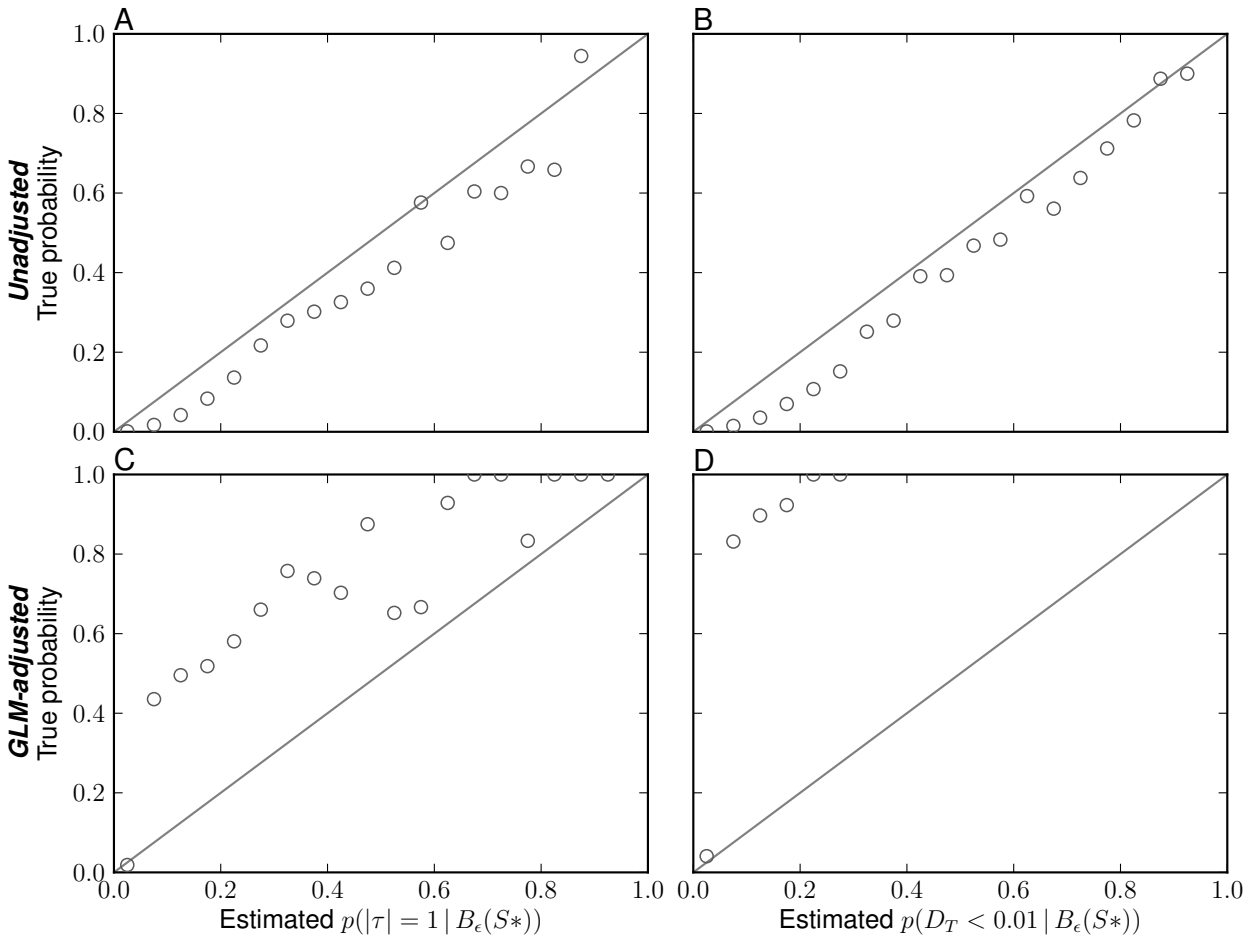

Figure 3.34. Model-choice accuracy for model M°_{DPP} when analyzing data generated under M°_{DPP}. The estimated posterior probability of a single divergence event, based on (A & C) $|\boldsymbol{\tau}| = 1$ and (B & D) $D_T < 0.01$, from 50,000 posterior estimates are assigned to bins of width 0.05 and plotted against the proportion of replicates in each bin where the truth is $|\boldsymbol{\tau}| = 1$ or $D_T < 0.01$. Results based on the (A & B) unadjusted and (C & D) GLM-adjusted posterior estimates are shown.

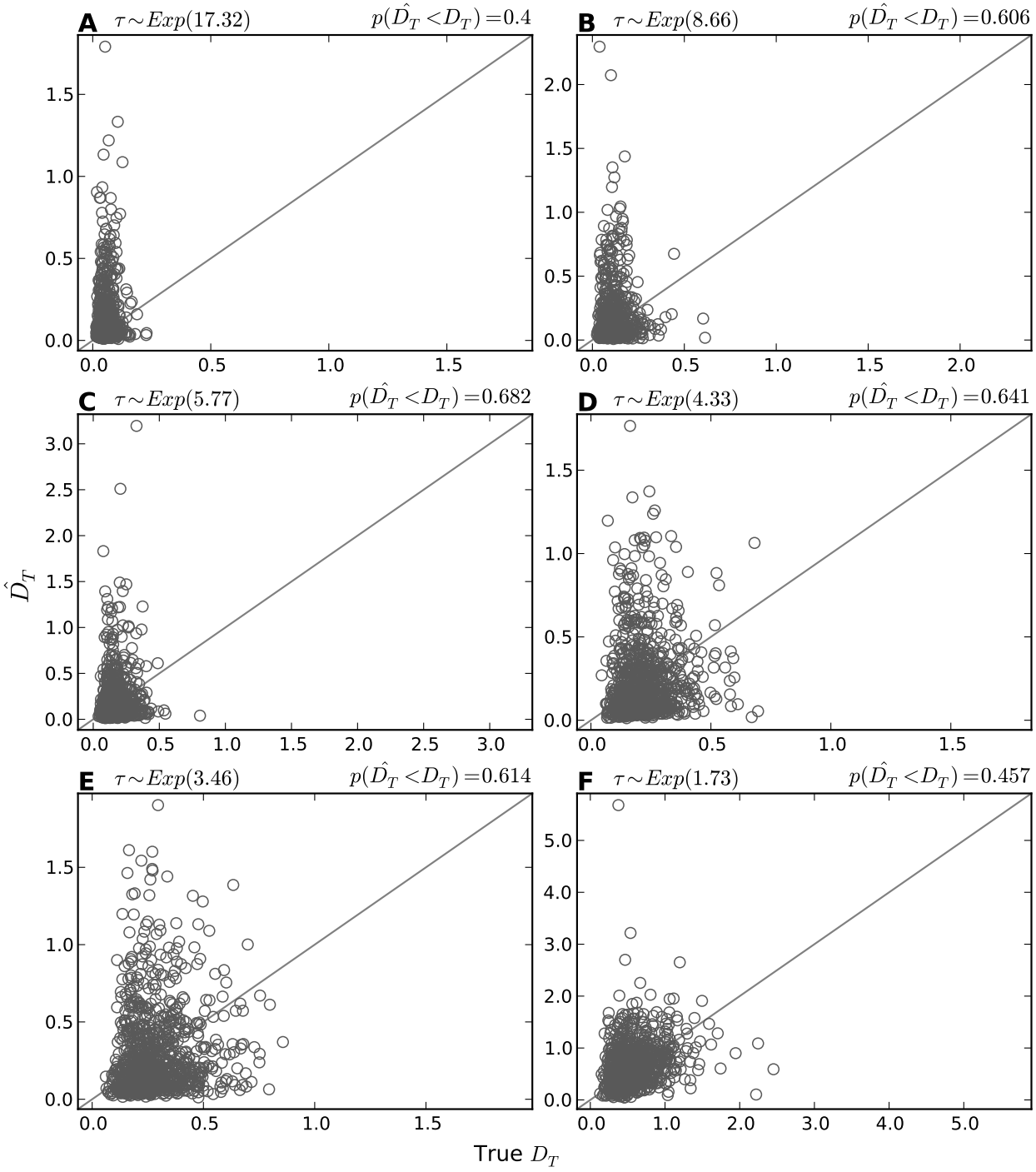

Figure 3.35. Estimation accuracy for model M_{DPP} when analyzing data generated under the series of models $\mathcal{M}_{Exp}$. The true versus estimated value of the dispersion index of divergence times (D_T) is plotted for 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, and the proportion of estimates less than the truth, $p(\hat{D}_T < D_T)$, is shown for each data model.

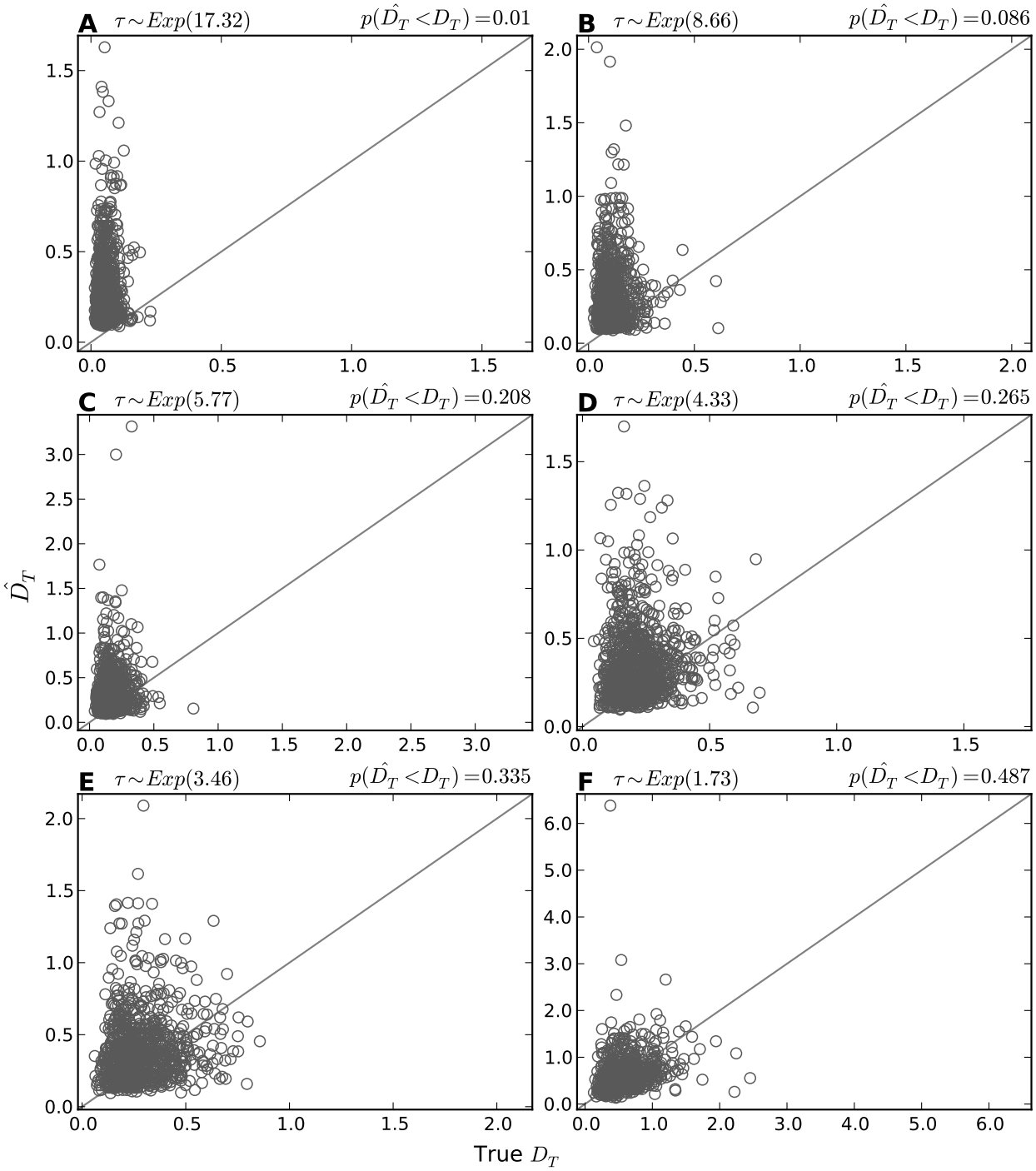

Figure 3.36. Estimation accuracy for model $M_{Uniform}$ when analyzing data generated under the series of models $\mathcal{M}_{Exp}$. The true versus estimated value of the dispersion index of divergence times (D_T) is plotted for 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, and the proportion of estimates less than the truth, $p(\hat{D_T} < D_T)$, is shown for each data model.

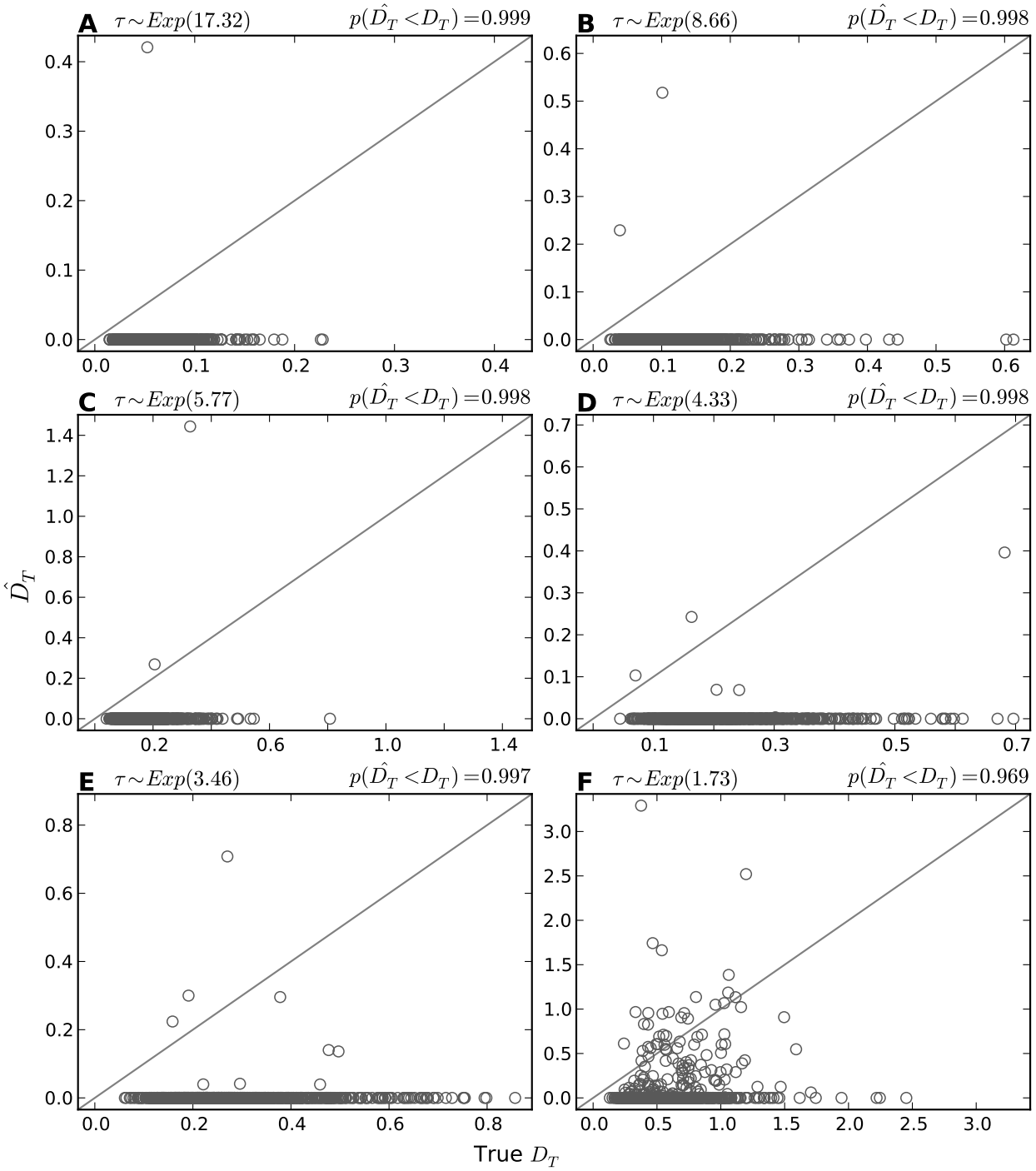

Figure 3.37. Estimation accuracy for model $M_{msBayes}$ when analyzing data generated under the series of models $\mathcal{M}_{Exp}$. The true versus estimated value of the dispersion index of divergence times (D_T) is plotted for 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, and the proportion of estimates less than the truth, $p(\hat{D}_T < D_T)$, is shown for each data model.

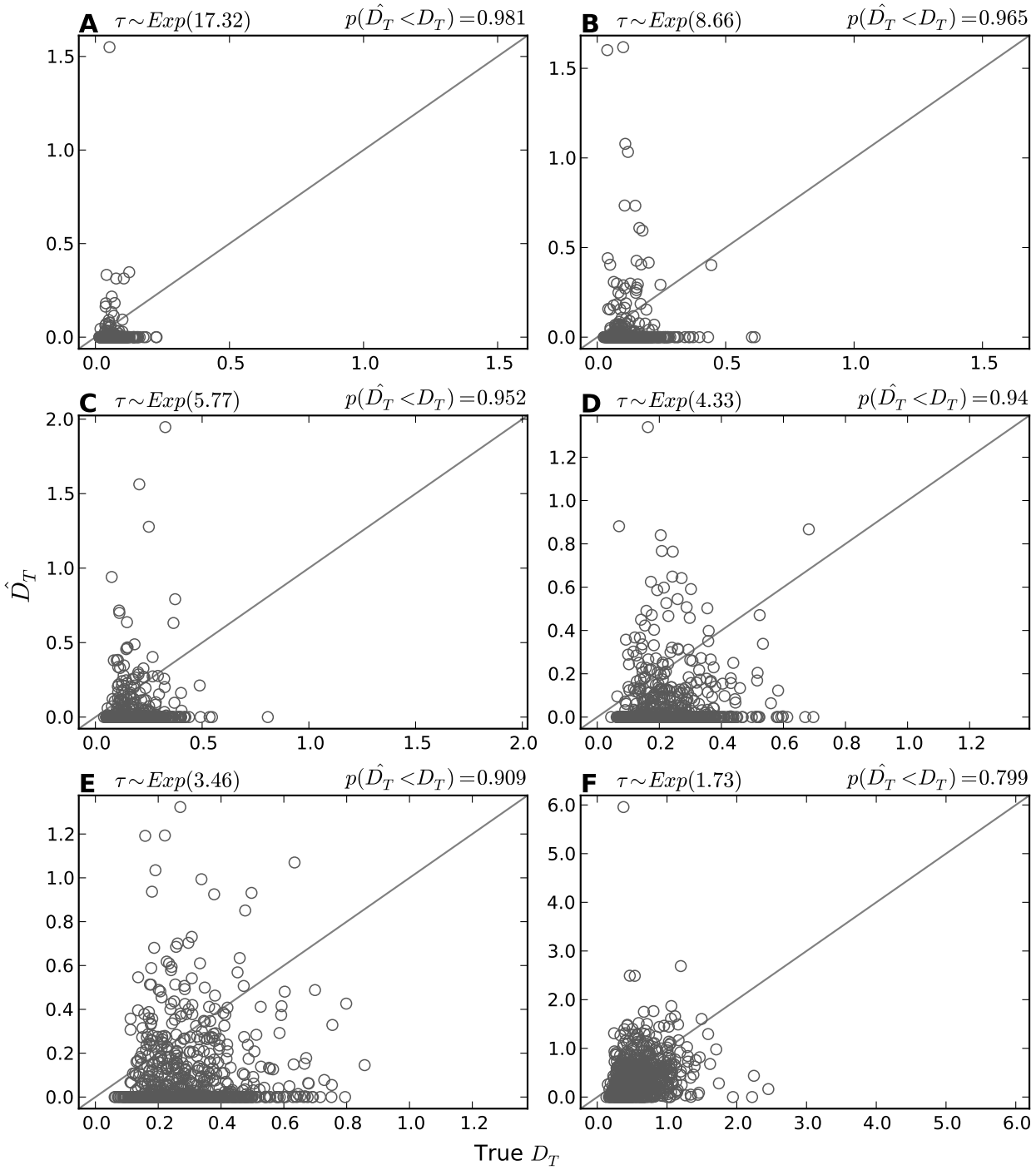

Figure 3.38. Estimation accuracy for model $M_{Ushaped}$ when analyzing data generated under the series of models $\mathcal{M}_{Exp}$. The true versus estimated value of the dispersion index of divergence times (D_T) is plotted for 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, and the proportion of estimates less than the truth, $p(\hat{D_T} < D_T)$, is shown for each data model.

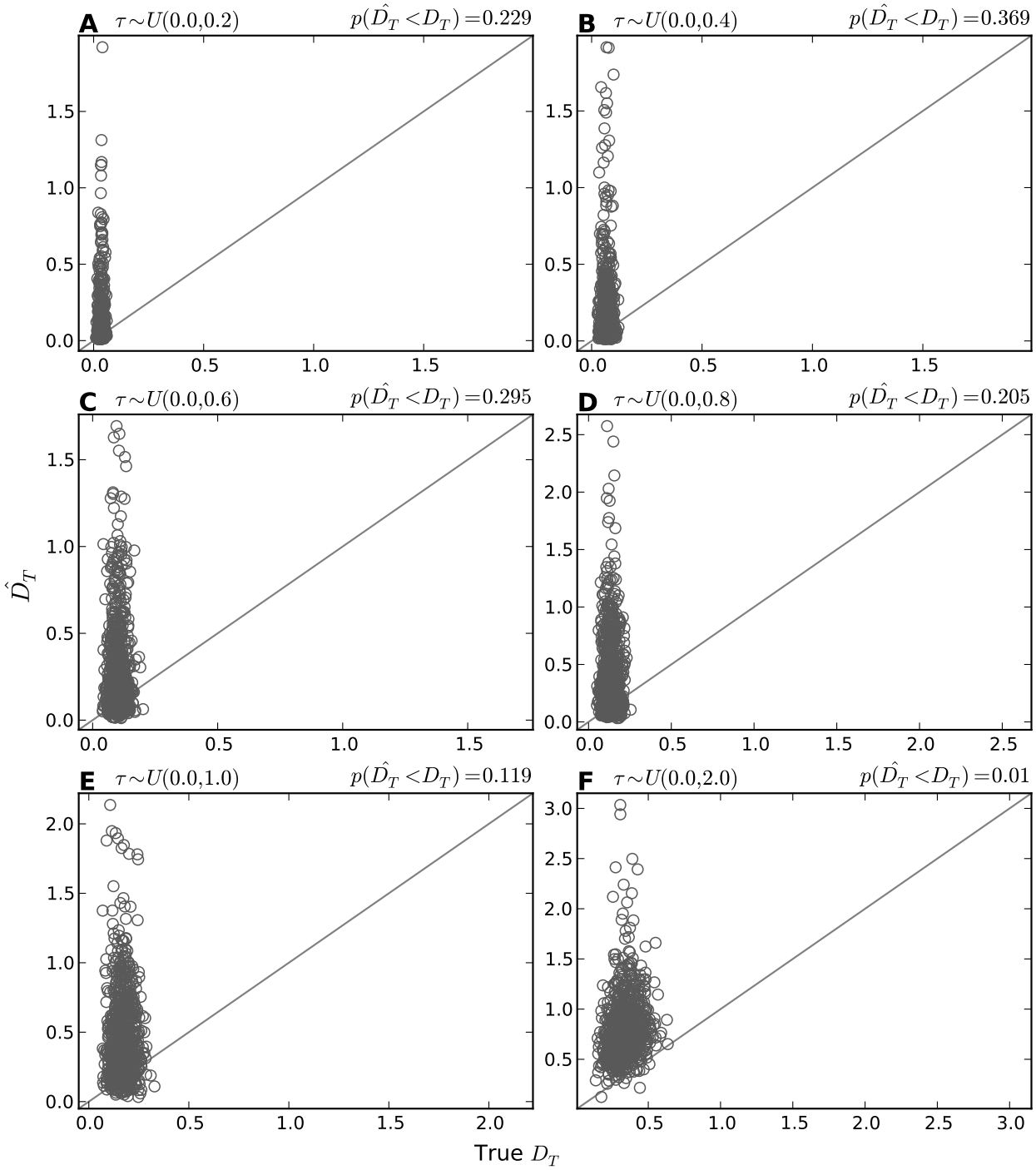

Figure 3.39. Estimation accuracy for model M_{DPP} when analyzing data generated under the series of models $\mathcal{M}_{Uniform}$. The true versus estimated value of the dispersion index of divergence times (D_T) is plotted for 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, and the proportion of estimates less than the truth, $p(\hat{D_T} < D_T)$, is shown for each data model.

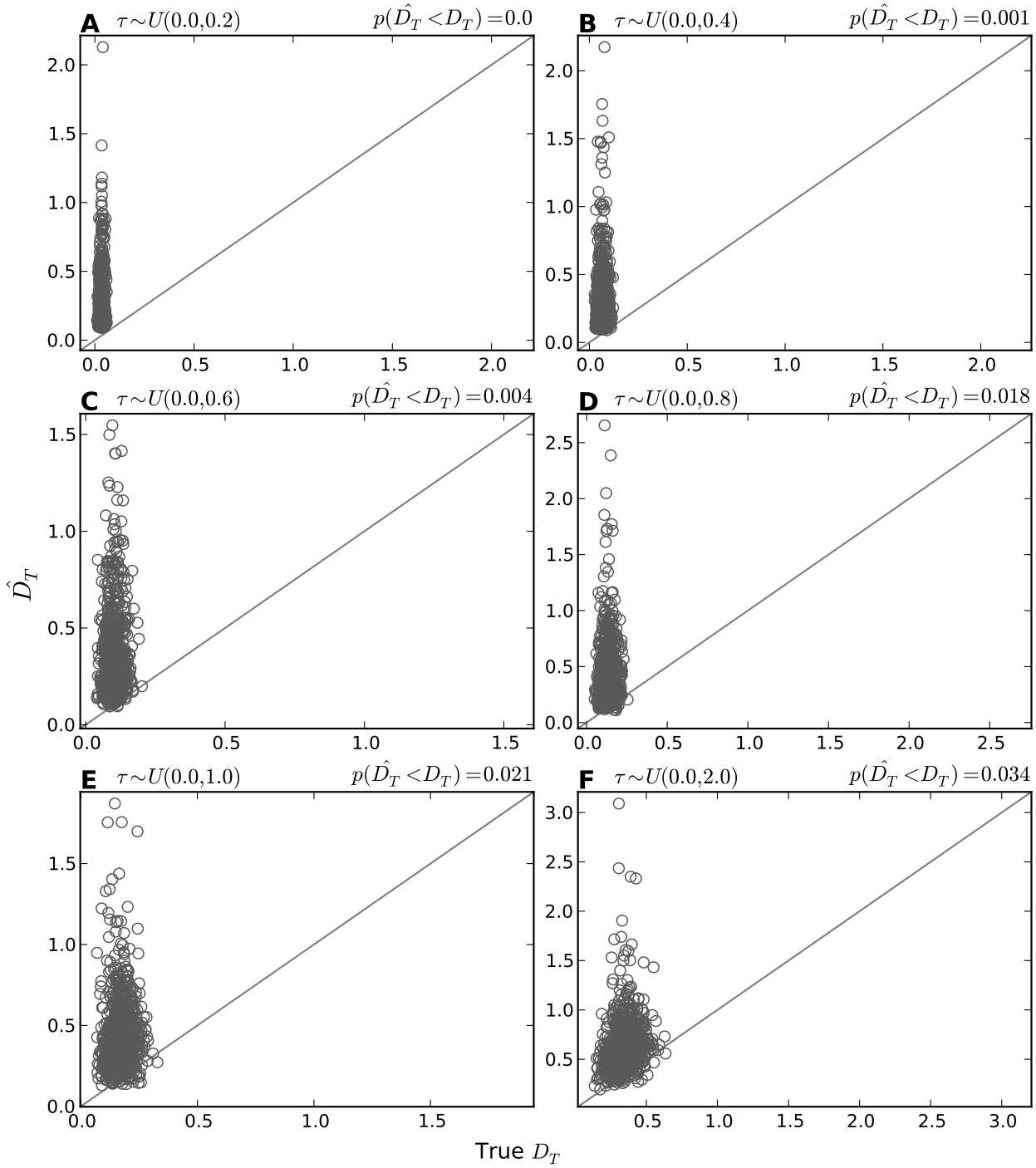

Figure 3.40. Estimation accuracy for model $M_{Uniform}$ when analyzing data generated under the series of models $\mathcal{M}_{Uniform}$. The true versus estimated value of the dispersion index of divergence times (D_T) is plotted for 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, and the proportion of estimates less than the truth, $p(\hat{D_T} < D_T)$, is shown for each data model.

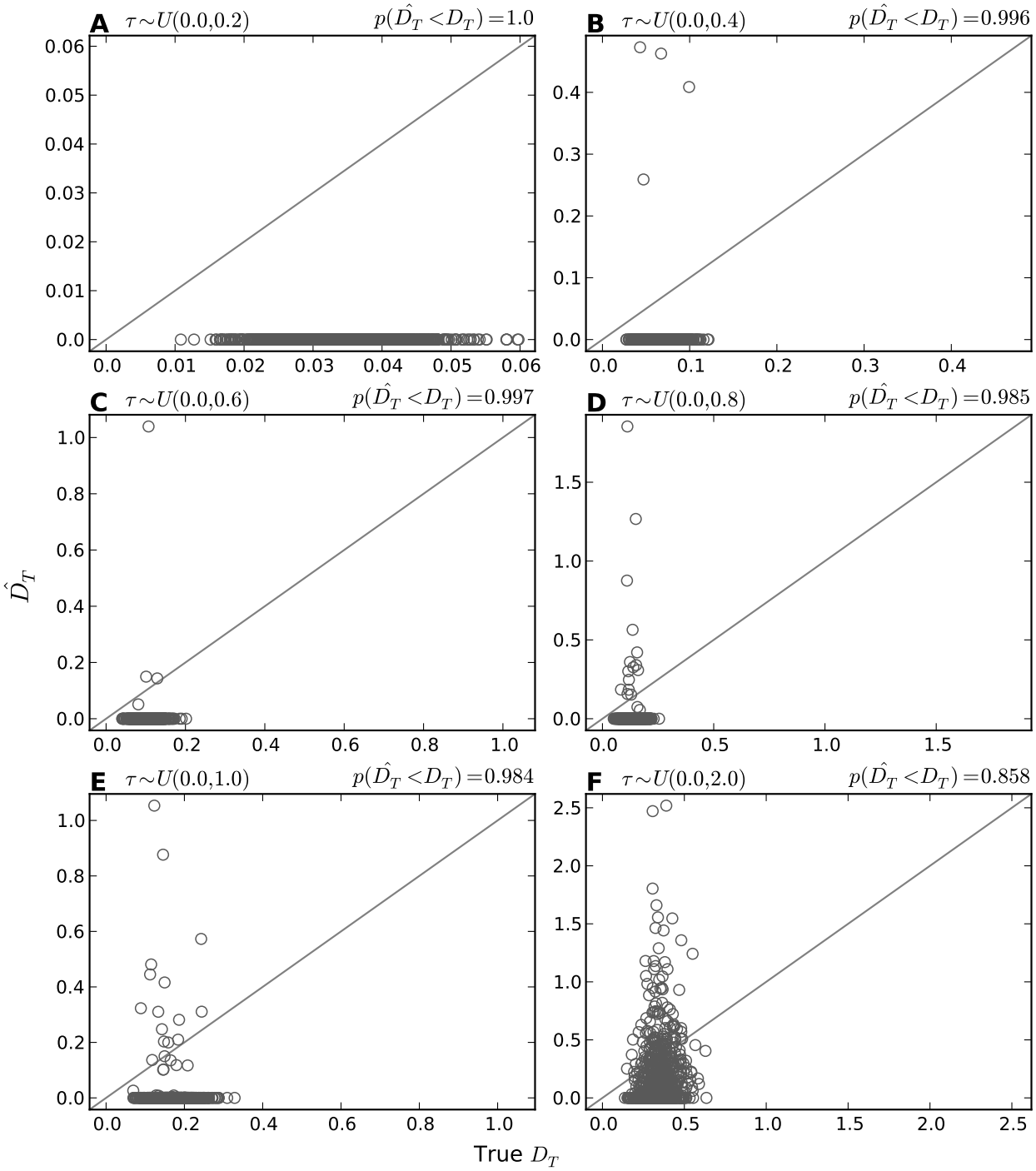

Figure 3.41. Estimation accuracy for model $M_{msBayes}$ when analyzing data generated under the series of models $\mathcal{M}_{Uniform}$. The true versus estimated value of the dispersion index of divergence times (D_T) is plotted for 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, and the proportion of estimates less than the truth, $p(\hat{D}_T < D_T)$, is shown for each data model.

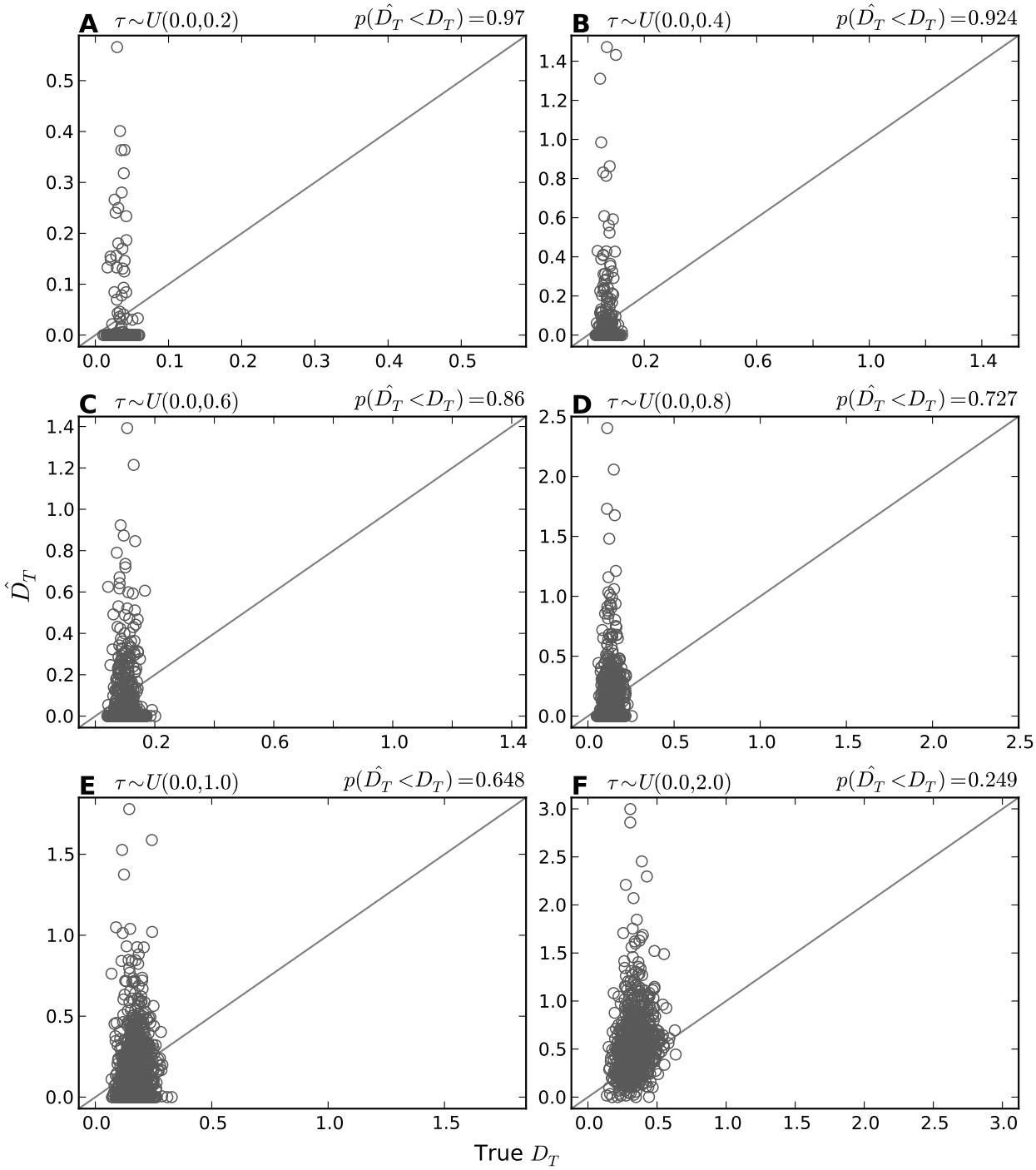

Figure 3.42. Estimation accuracy for model $M_{Ushaped}$ when analyzing data generated under the series of models $\mathcal{M}_{Uniform}$. The true versus estimated value of the dispersion index of divergence times (D_T) is plotted for 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, and the proportion of estimates less than the truth, $p(\hat{D}_T < D_T)$, is shown for each data model.

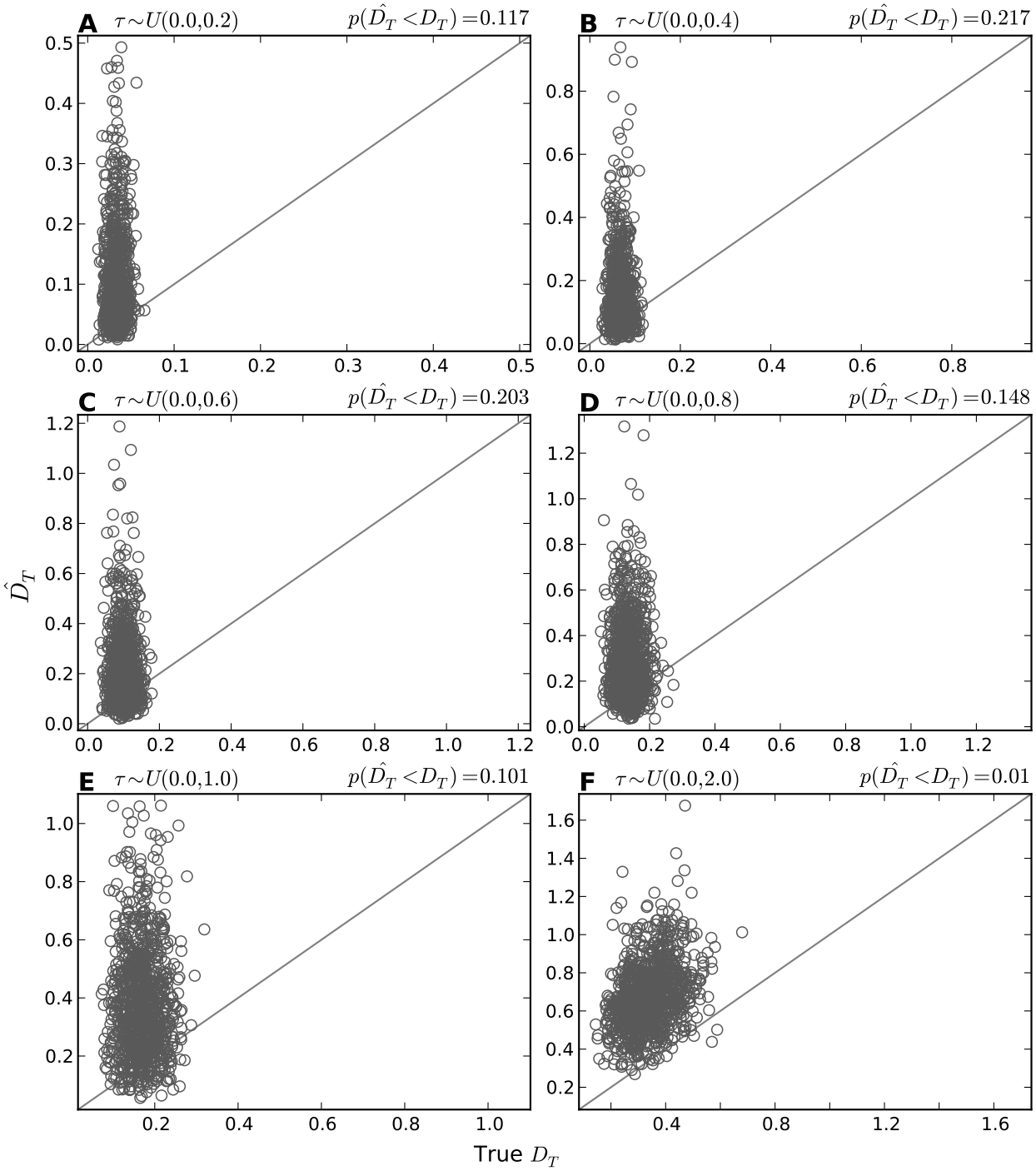

Figure 3.43. Estimation accuracy for model M_{DPP} when analyzing data generated under the series of models $\mathcal{M}_{msBayes}$. The true versus estimated value of the dispersion index of divergence times (D_T) is plotted for 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, and the proportion of estimates less than the truth, $p(\hat{D_T} < D_T)$, is shown for each data model.

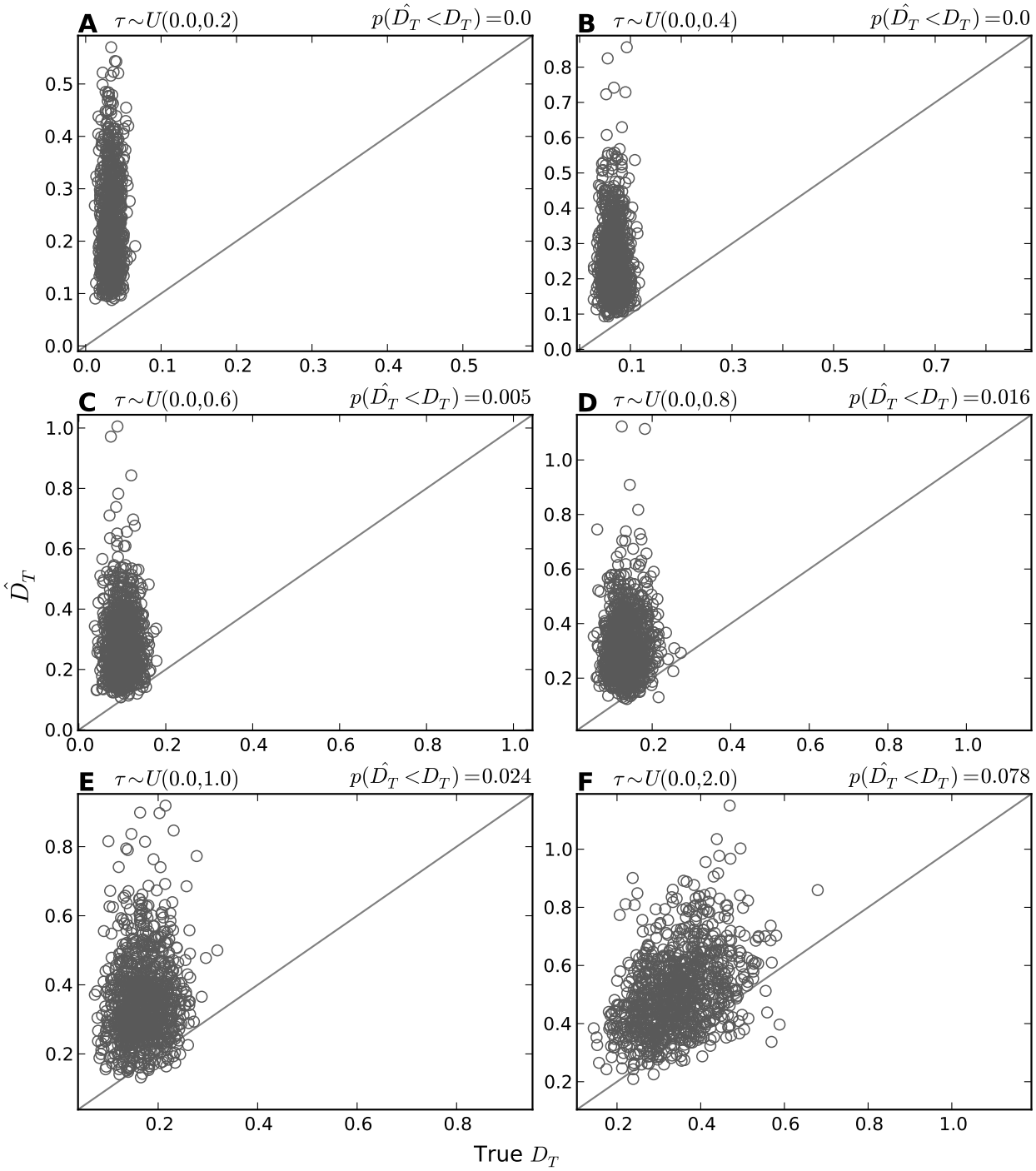

Figure 3.44. Estimation accuracy for model $M_{Uniform}$ when analyzing data generated under the series of models $\mathcal{M}_{msBayes}$. The true versus estimated value of the dispersion index of divergence times (D_T) is plotted for 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, and the proportion of estimates less than the truth, $p(\hat{D}_T < D_T)$, is shown for each data model.

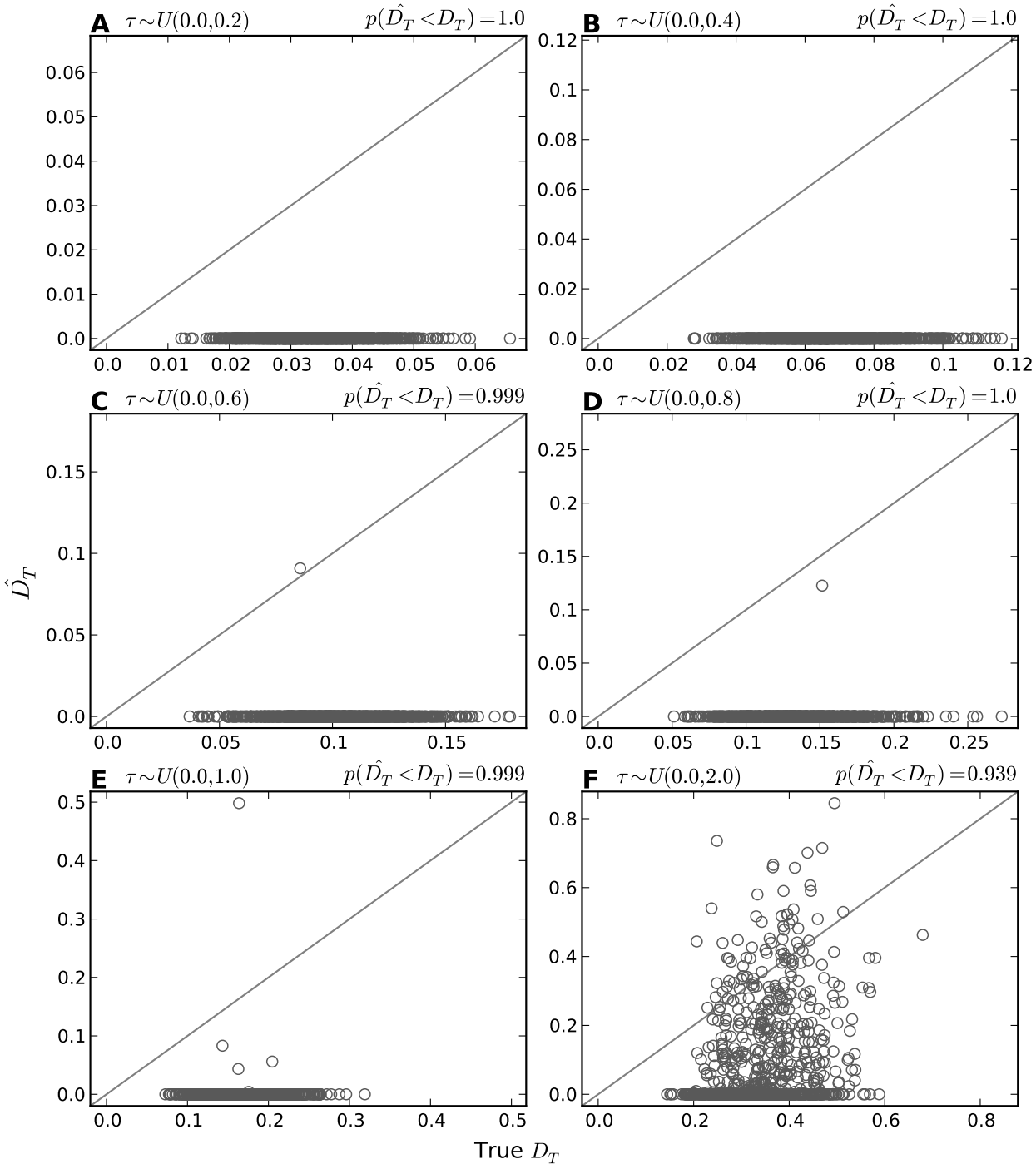

Figure 3.45. Estimation accuracy for model $M_{msBayes}$ when analyzing data generated under the series of models $\mathcal{M}_{msBayes}$. The true versus estimated value of the dispersion index of divergence times (D_T) is plotted for 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, and the proportion of estimates less than the truth, $p(\hat{D}_T < D_T)$, is shown for each data model.

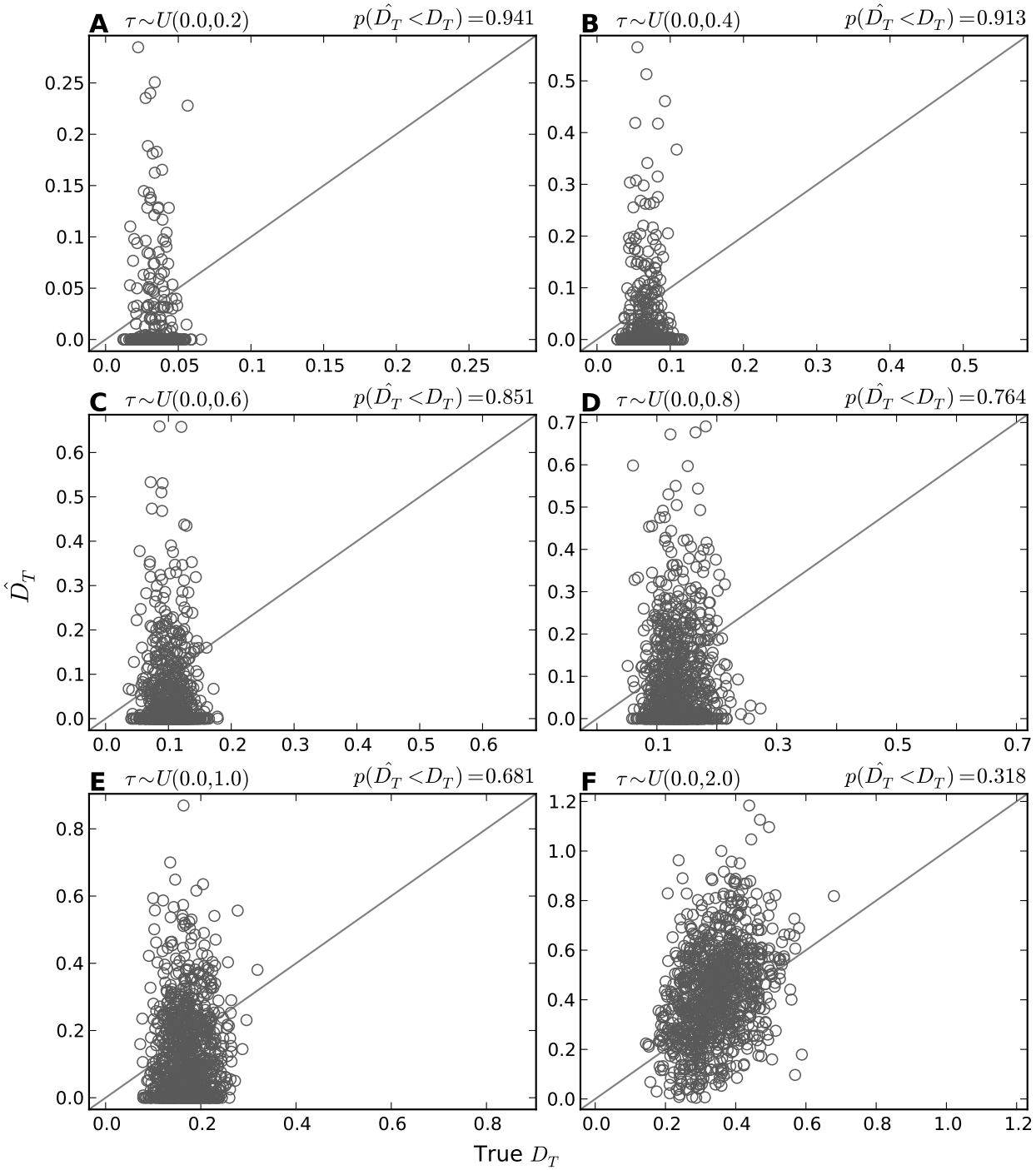

Figure 3.46. Estimation accuracy for model $M_{Ushaped}$ when analyzing data generated under the series of models $\mathcal{M}_{msBayes}$. The true versus estimated value of the dispersion index of divergence times (D_T) is plotted for 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, and the proportion of estimates less than the truth, $p(\hat{D}_T < D_T)$, is shown for each data model.

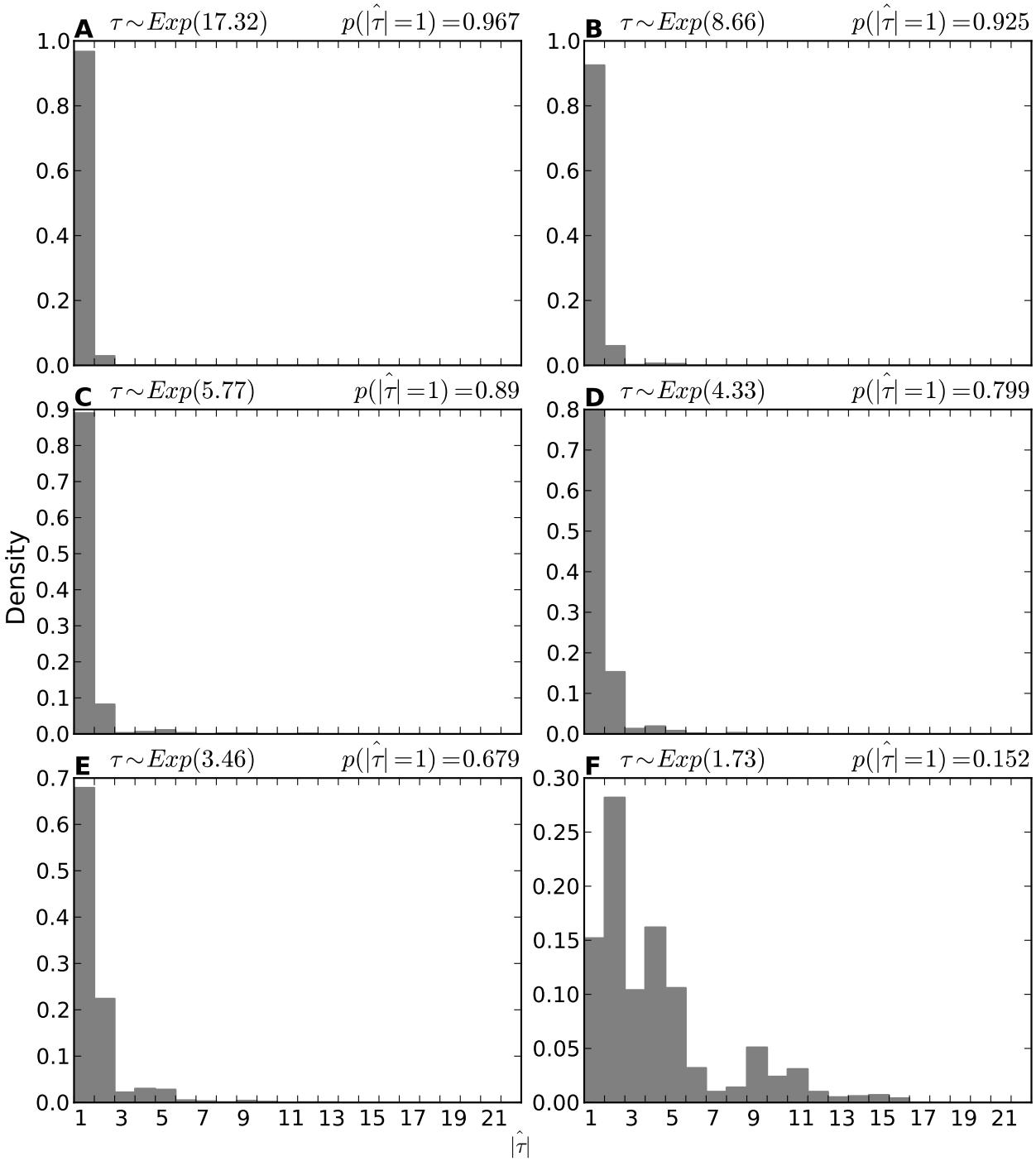

Figure 3.47. The power of model M_{DPP} to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate the estimated number of divergence events ($|\hat{\boldsymbol{\tau}}|$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model inferring one divergence event, $p(|\hat{\boldsymbol{\tau}}| = 1)$, given for each data model.

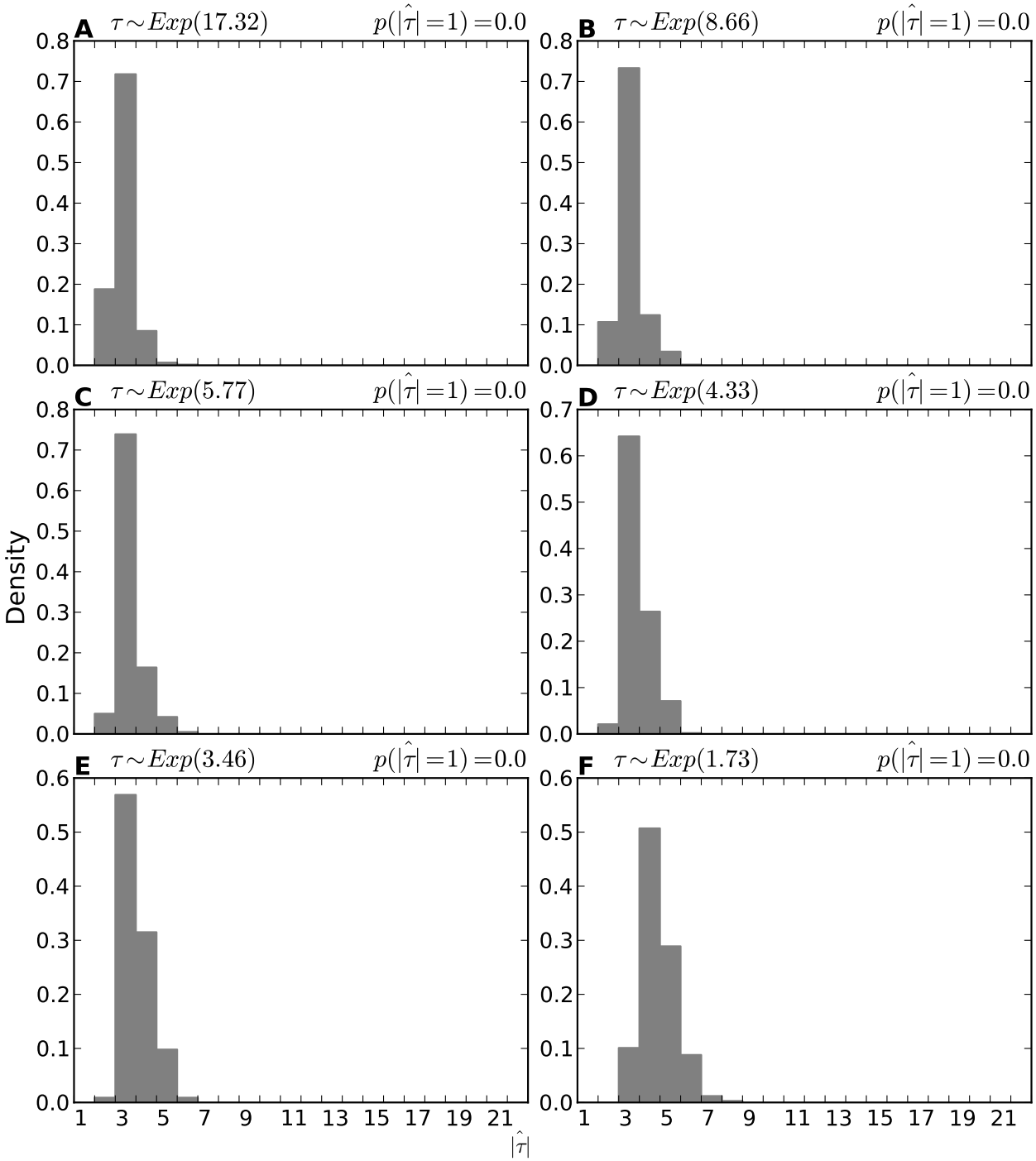

Figure 3.48. The power of model $M_{Uniform}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate the estimated number of divergence events ($|\hat{\boldsymbol{\tau}}|$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model inferring one divergence event, $p(|\hat{\boldsymbol{\tau}}| = 1)$, given for each data model.

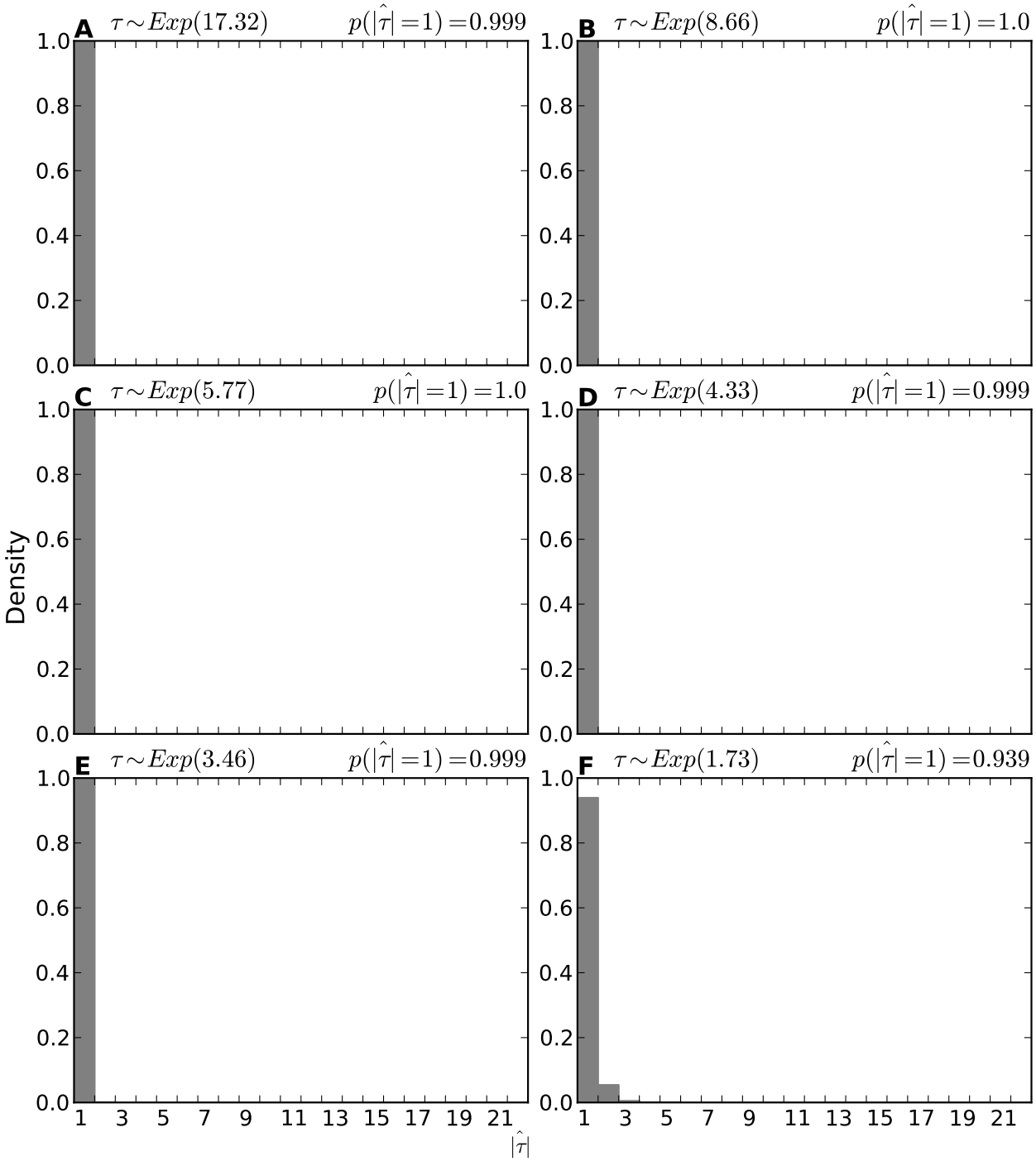

Figure 3.49. The power of model $M_{msBayes}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate the estimated number of divergence events ($|\hat{\boldsymbol{\tau}}|$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model inferring one divergence event, $p(|\hat{\boldsymbol{\tau}}| = 1)$, given for each data model.

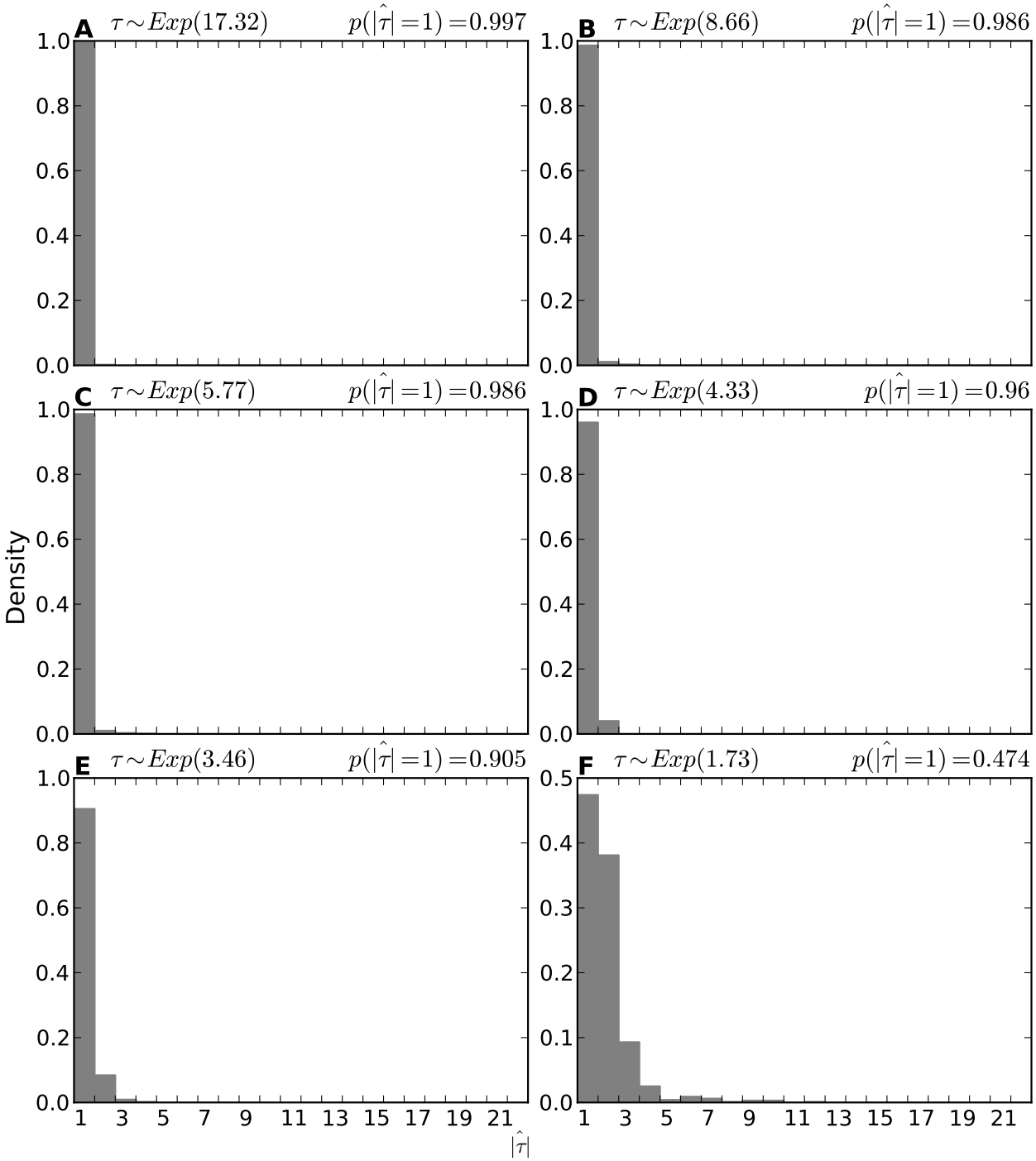

Figure 3.50. The power of model $M_{Ushaped}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate the estimated number of divergence events ($|\hat{\boldsymbol{\tau}}|$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model inferring one divergence event, $p(|\hat{\boldsymbol{\tau}}| = 1)$, given for each data model.

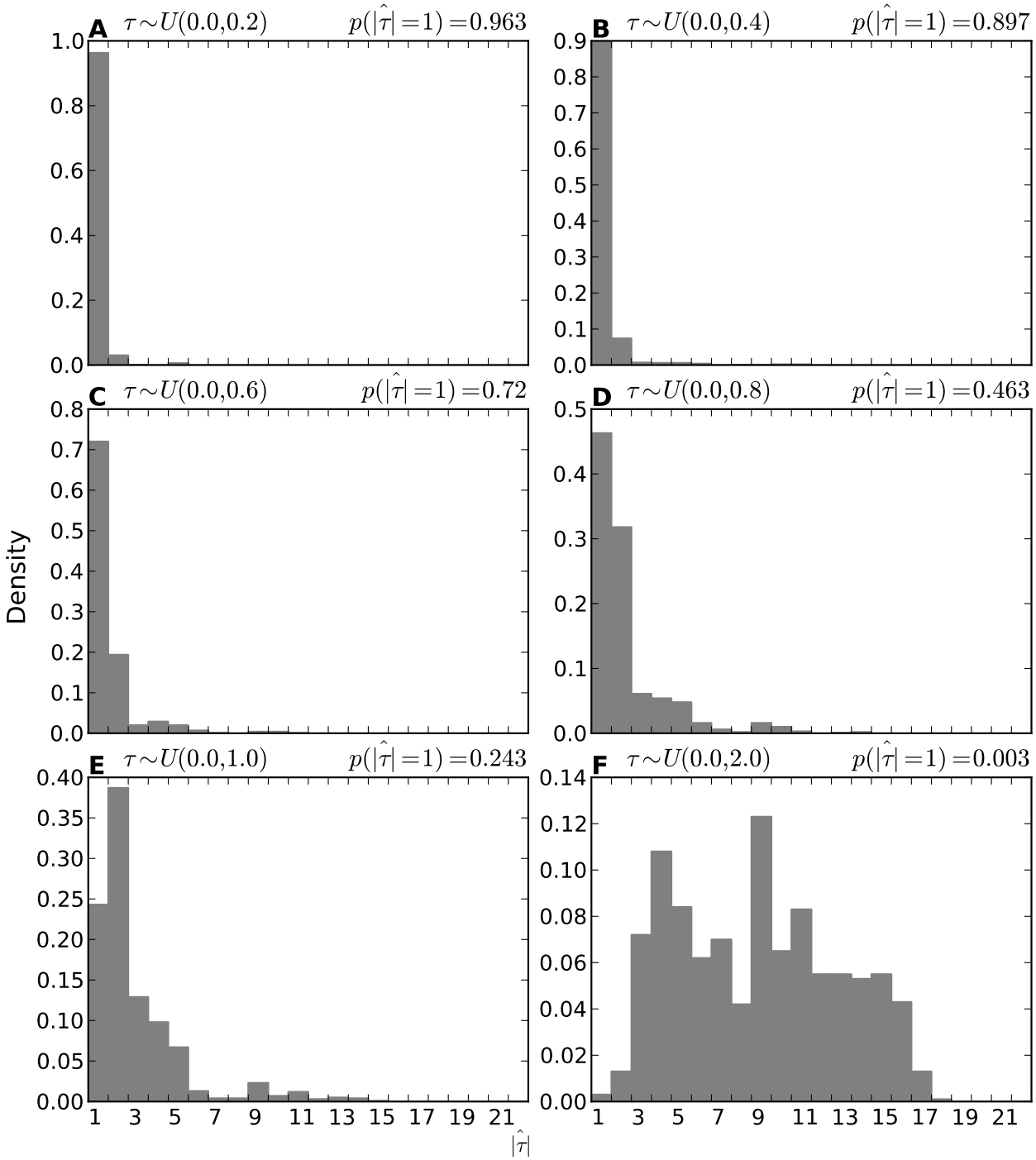

Figure 3.51. The power of model M_{DPP} to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate the estimated number of divergence events ($|\hat{\boldsymbol{\tau}}|$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model inferring one divergence event, $p(|\hat{\boldsymbol{\tau}}| = 1)$, given for each data model.

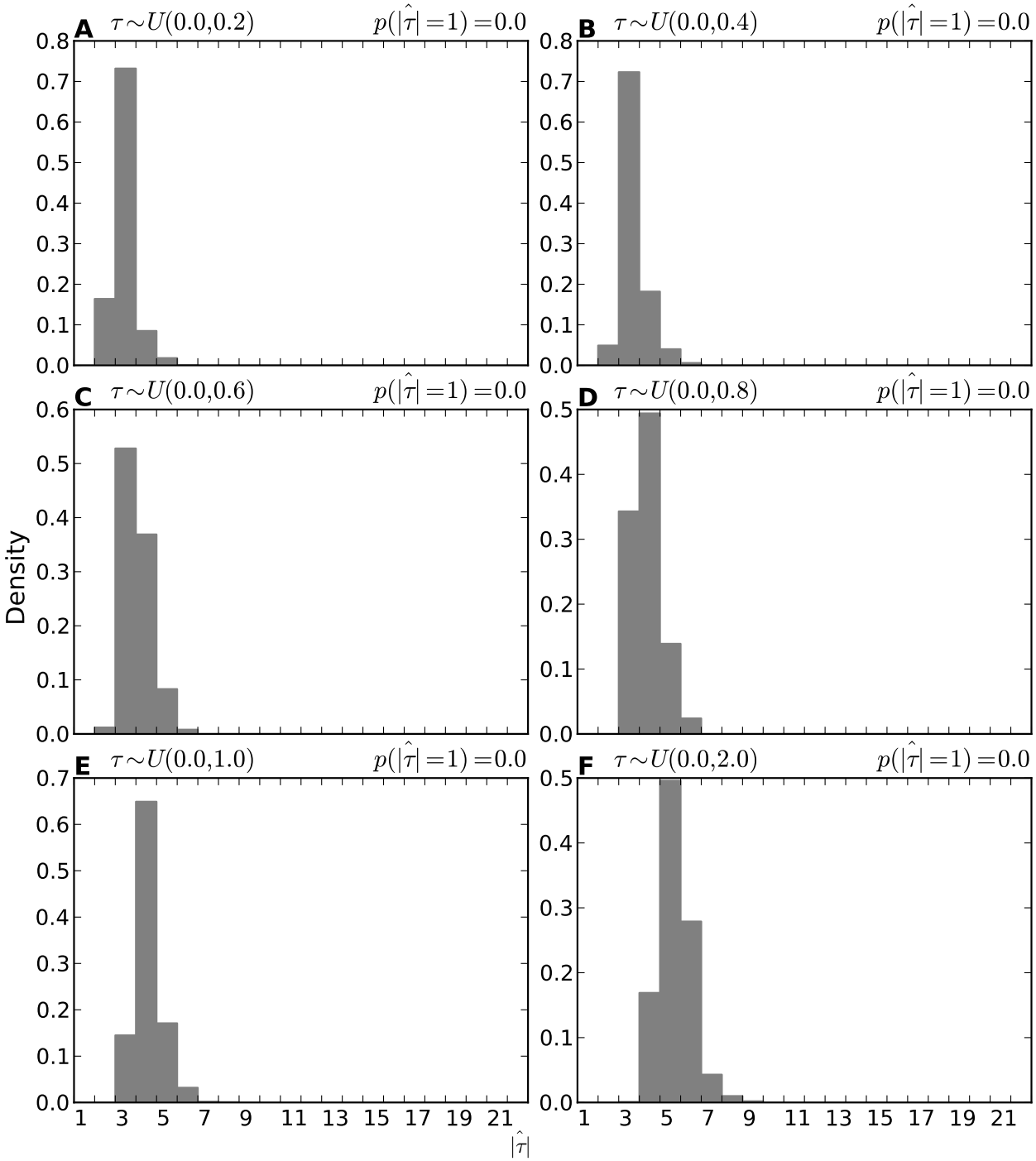

Figure 3.52. The power of model $M_{Uniform}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate the estimated number of divergence events ($|\hat{\boldsymbol{\tau}}|$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model inferring one divergence event, $p(|\hat{\boldsymbol{\tau}}| = 1)$, given for each data model.

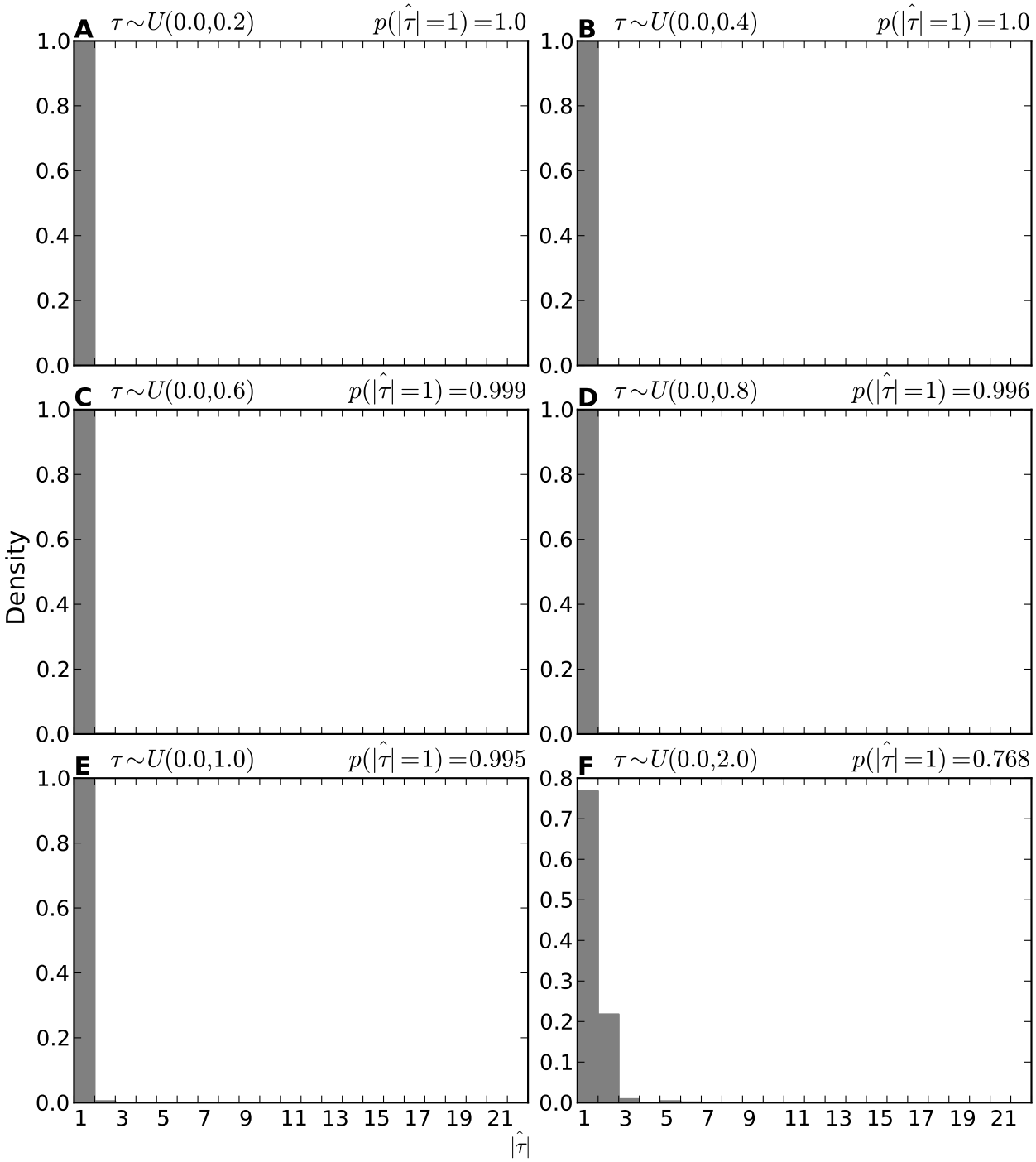

Figure 3.53. The power of model $M_{msBayes}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate the estimated number of divergence events ($|\hat{\boldsymbol{\tau}}|$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model inferring one divergence event, $p(|\hat{\boldsymbol{\tau}}| = 1)$, given for each data model.

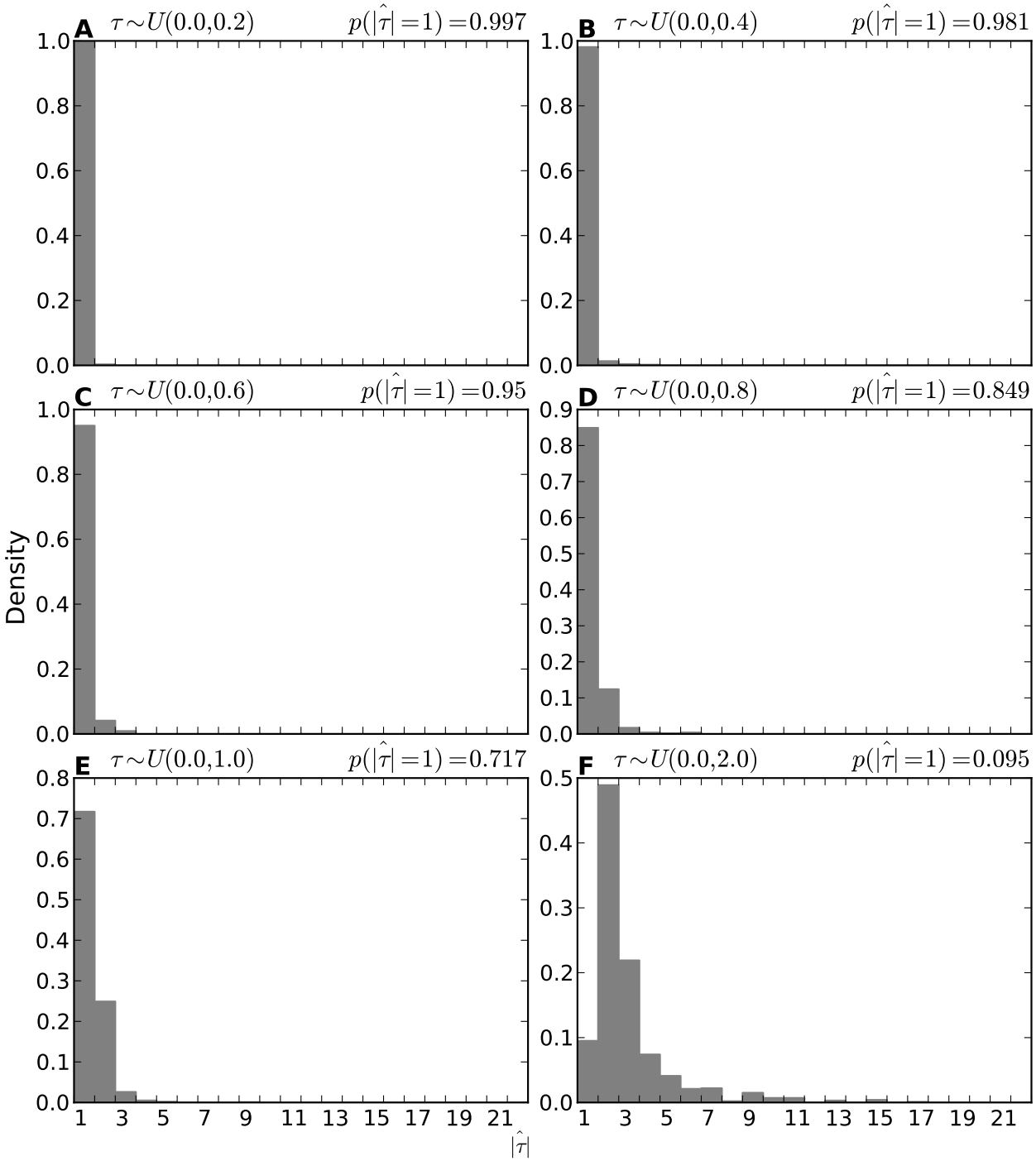

Figure 3.54. The power of model $M_{Ushaped}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate the estimated number of divergence events ($|\hat{\boldsymbol{\tau}}|$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model inferring one divergence event, $p(|\hat{\boldsymbol{\tau}}| = 1)$, given for each data model.

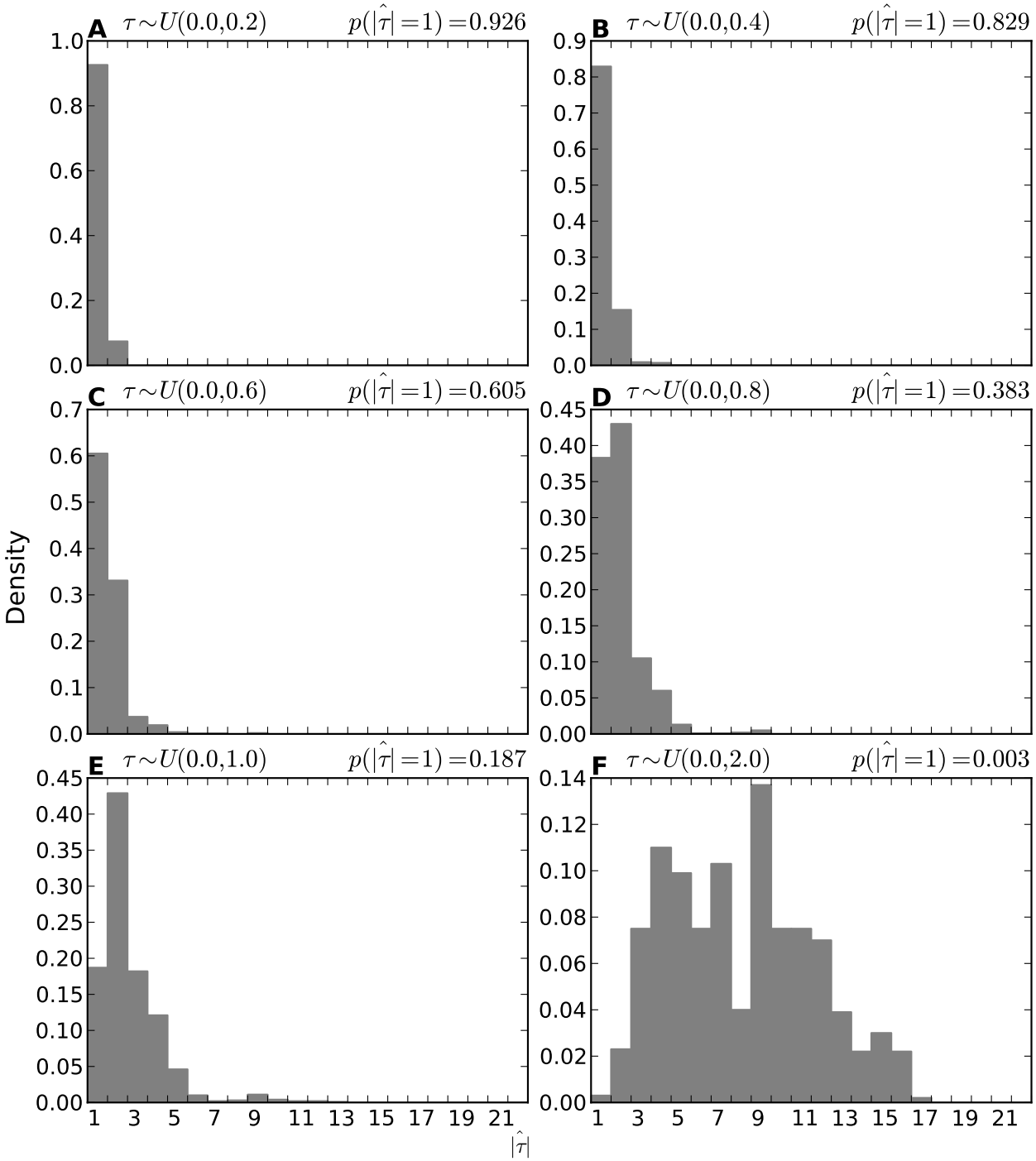

Figure 3.55. The power of model M_{DPP} to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate the estimated number of divergence events ($|\hat{\boldsymbol{\tau}}|$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model inferring one divergence event, $p(|\hat{\boldsymbol{\tau}}| = 1)$, given for each data model.

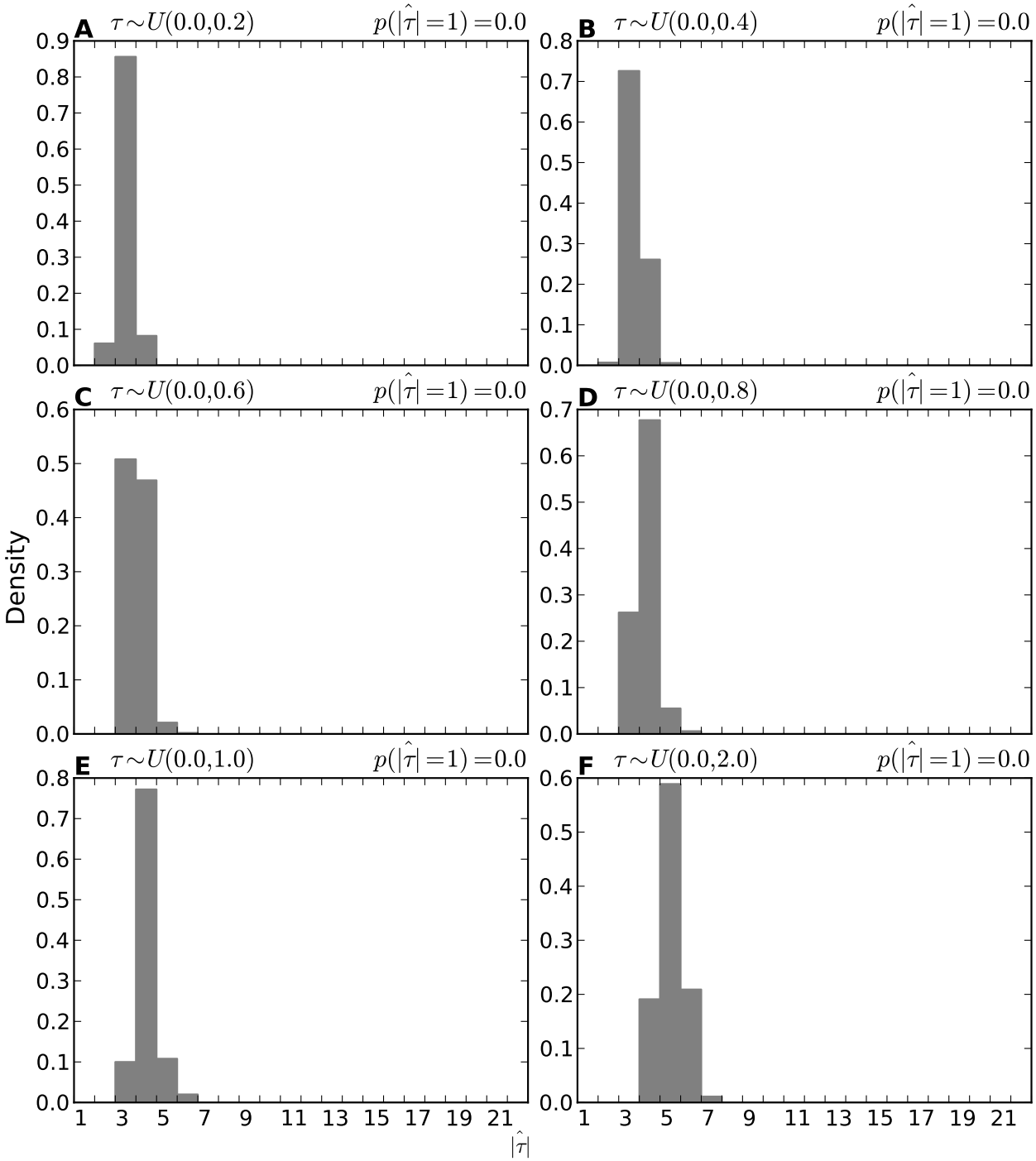

Figure 3.56. The power of model $M_{Uniform}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate the estimated number of divergence events ($|\hat{\boldsymbol{\tau}}|$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model inferring one divergence event, $p(|\hat{\boldsymbol{\tau}}| = 1)$, given for each data model.

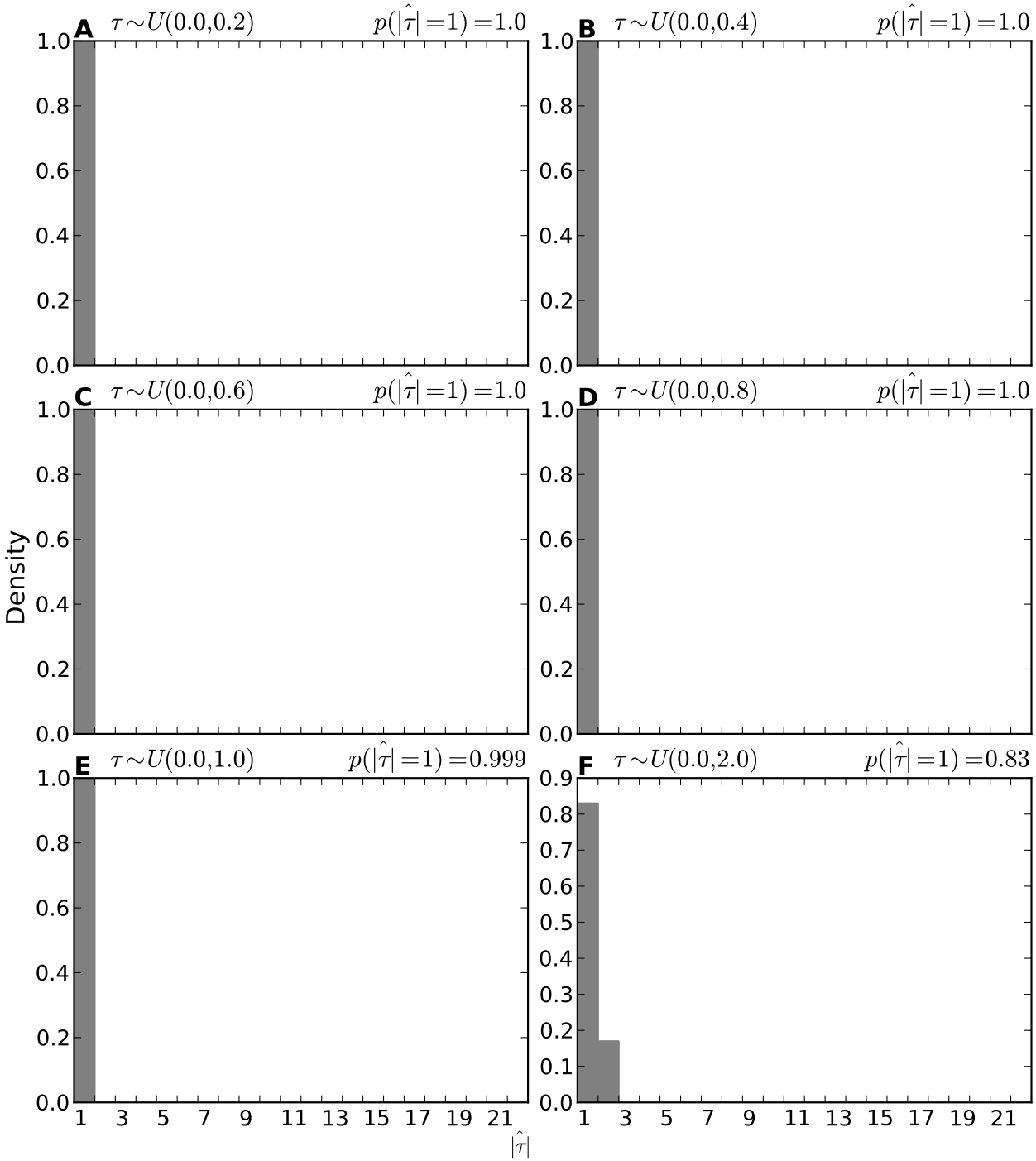

Figure 3.57. The power of model $M_{msBayes}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate the estimated number of divergence events ($|\hat{\boldsymbol{\tau}}|$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model inferring one divergence event, $p(|\hat{\boldsymbol{\tau}}| = 1)$, given for each data model.

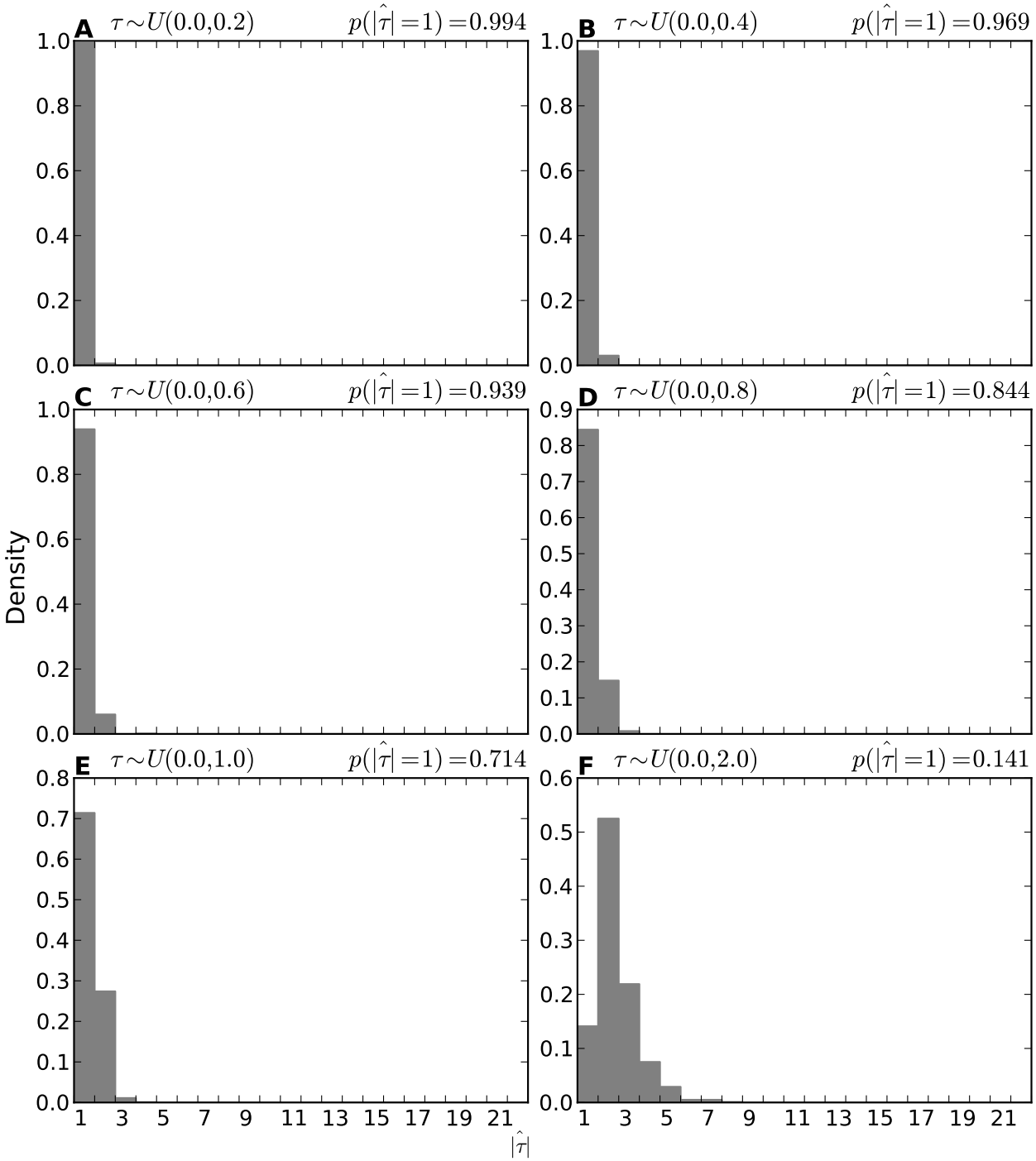

Figure 3.58. The power of model $M_{Ushaped}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate the estimated number of divergence events ($|\hat{\boldsymbol{\tau}}|$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model inferring one divergence event, $p(|\hat{\boldsymbol{\tau}}| = 1)$, given for each data model.

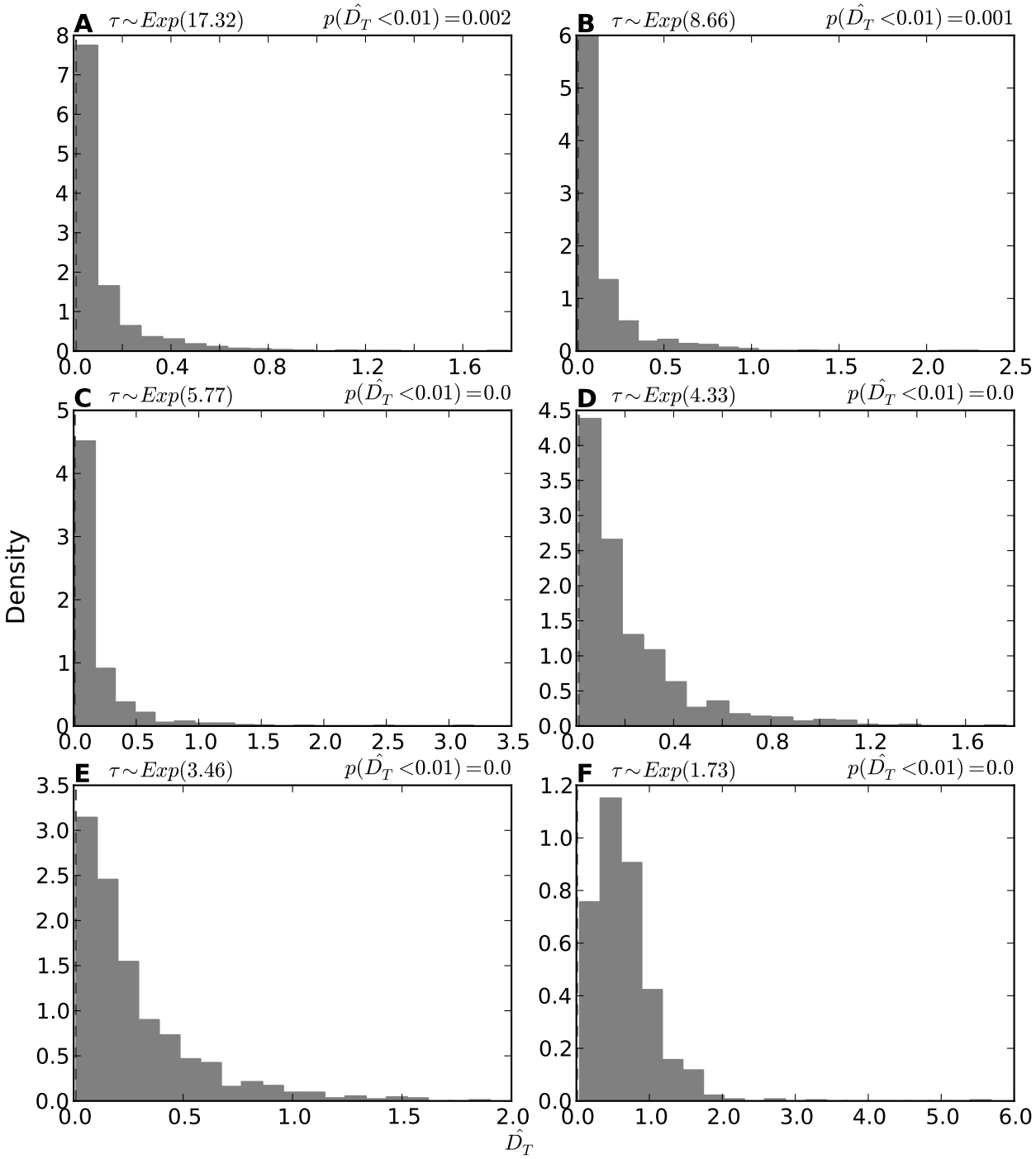

Figure 3.59. The power of model M_{DPP} to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate the estimated dispersion index of divergence times ($\hat{D_T}$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model inferring one divergence event, $p(\hat{D_T} < 0.01)$, given for each data model.

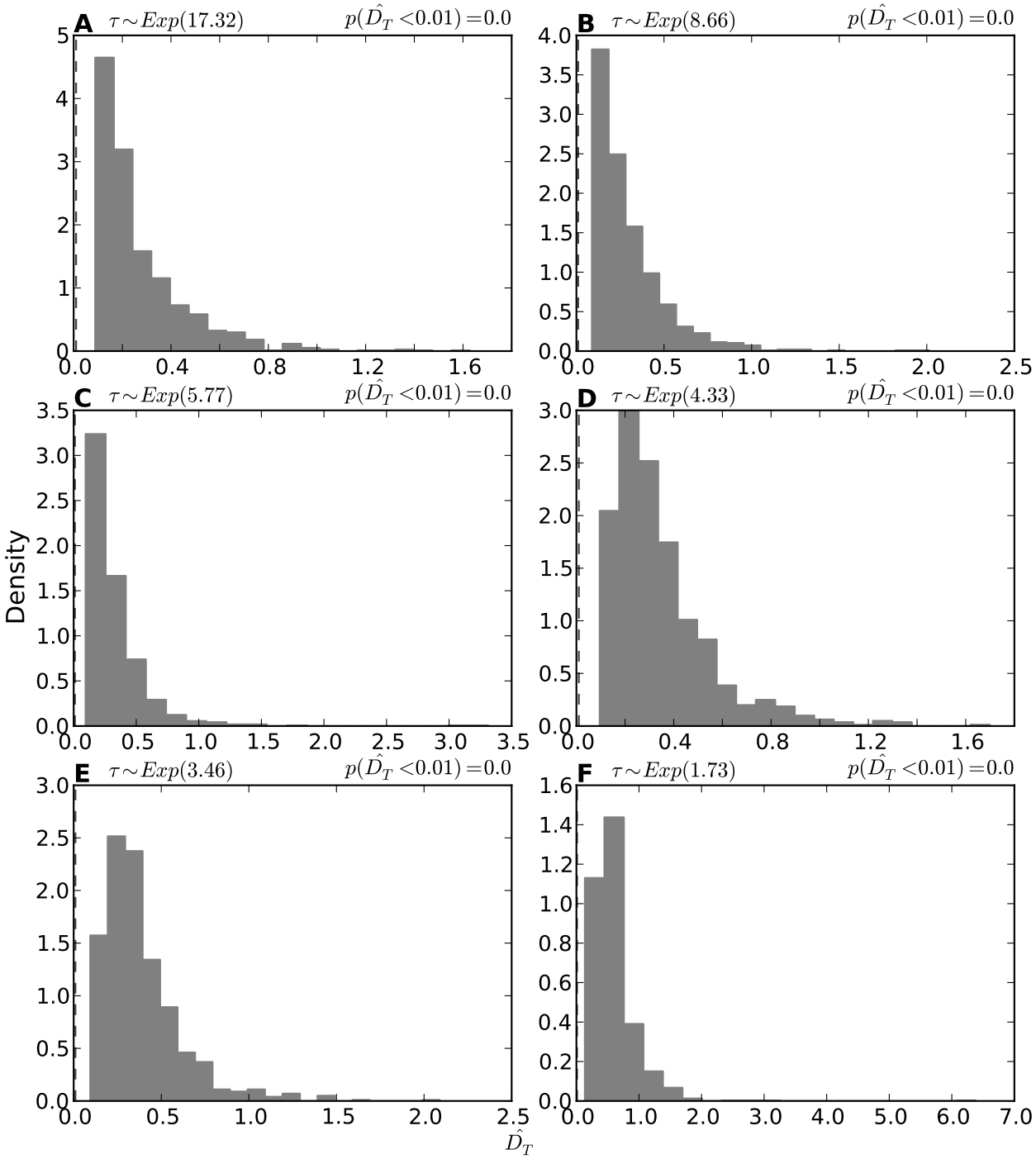

Figure 3.60. The power of model $M_{Uniform}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate the estimated dispersion index of divergence times ($\hat{D_T}$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model inferring one divergence event, $p(\hat{D_T} < 0.01)$, given for each data model.

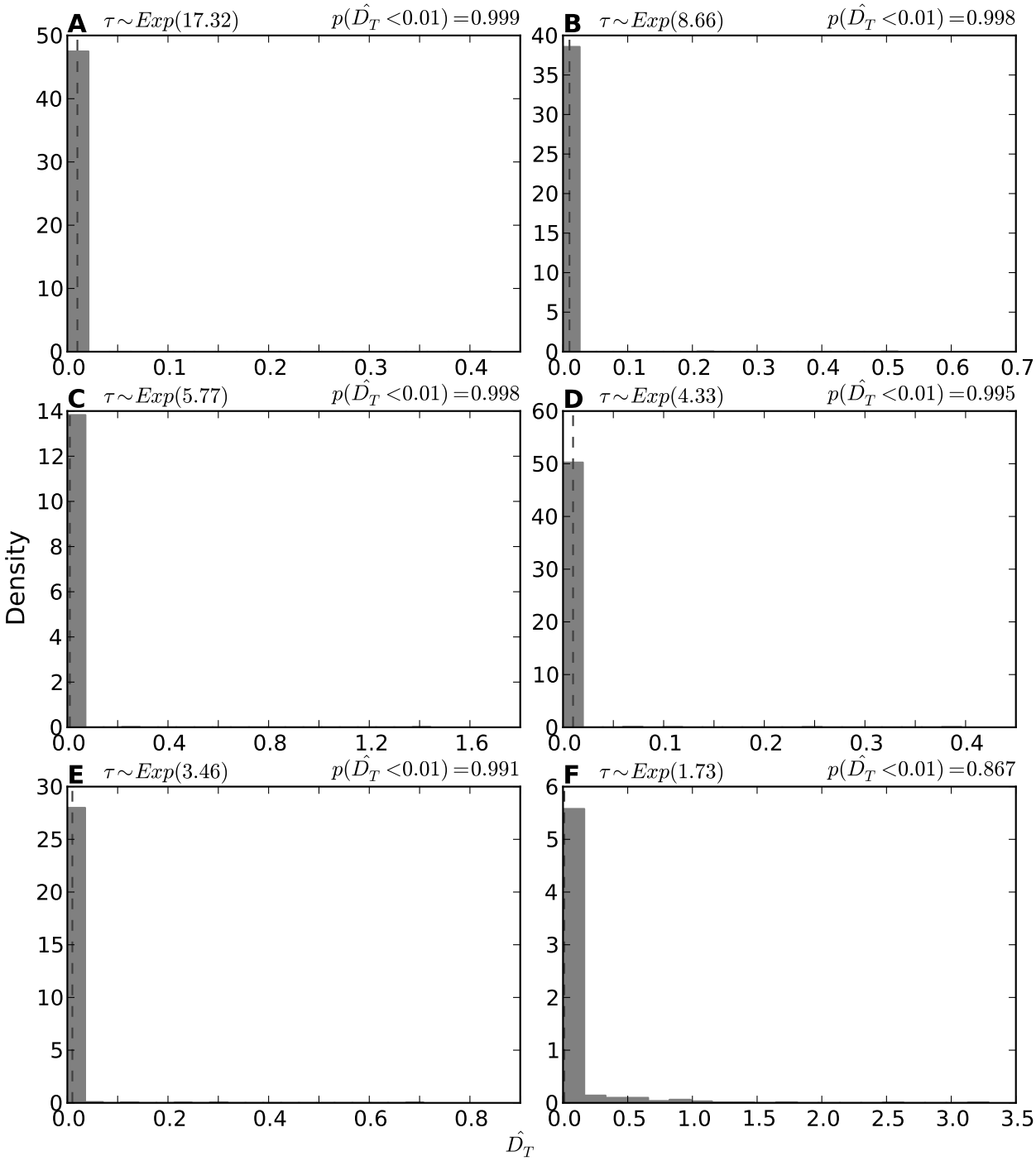

Figure 3.61. The power of model $M_{msBayes}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate the estimated dispersion index of divergence times ($\hat{D_T}$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model inferring one divergence event, $p(\hat{D_T} < 0.01)$, given for each data model.

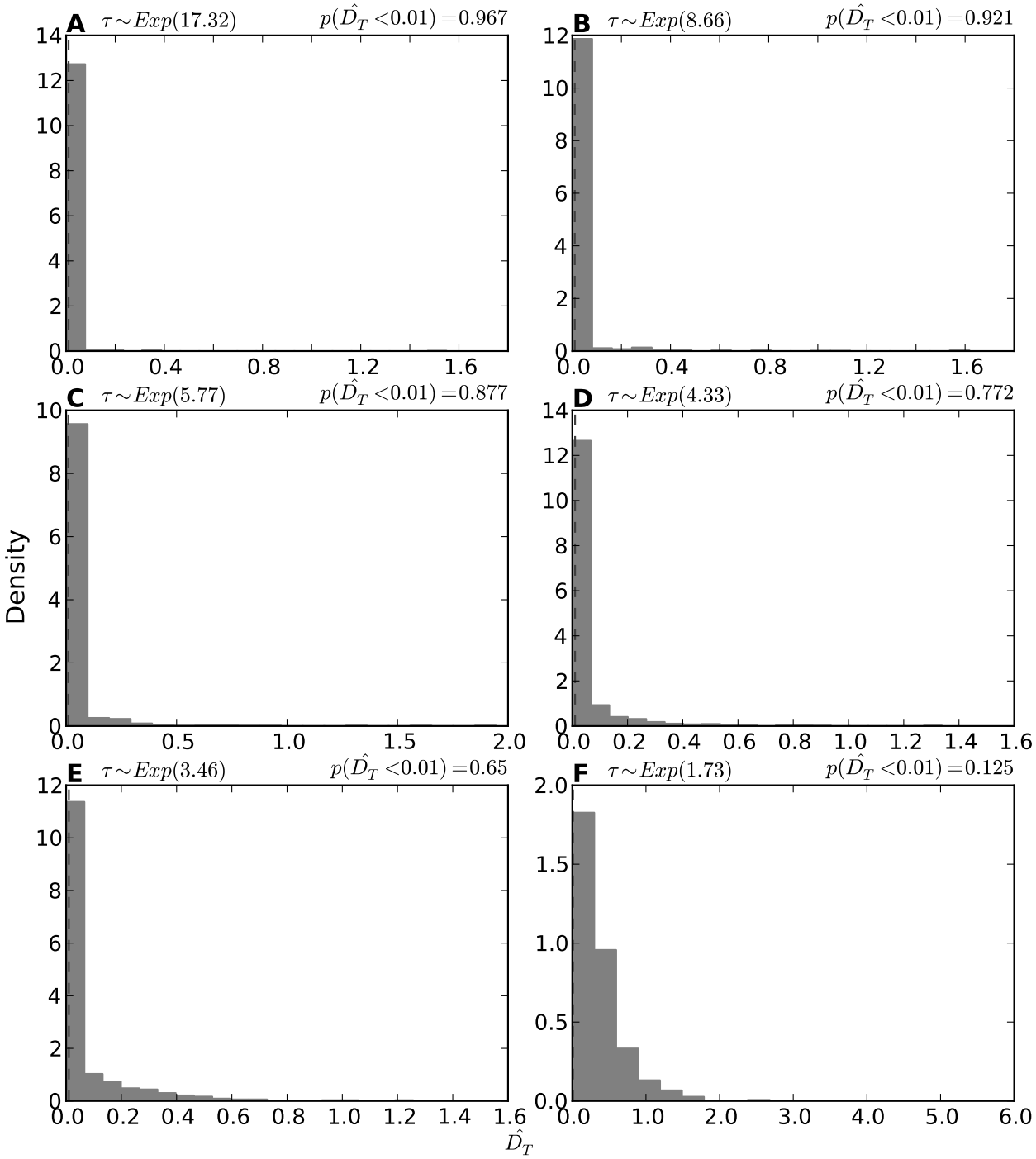

Figure 3.62. The power of model $M_{Ushaped}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate the estimated dispersion index of divergence times ($\hat{D_T}$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model inferring one divergence event, $p(\hat{D_T} < 0.01)$, given for each data model.

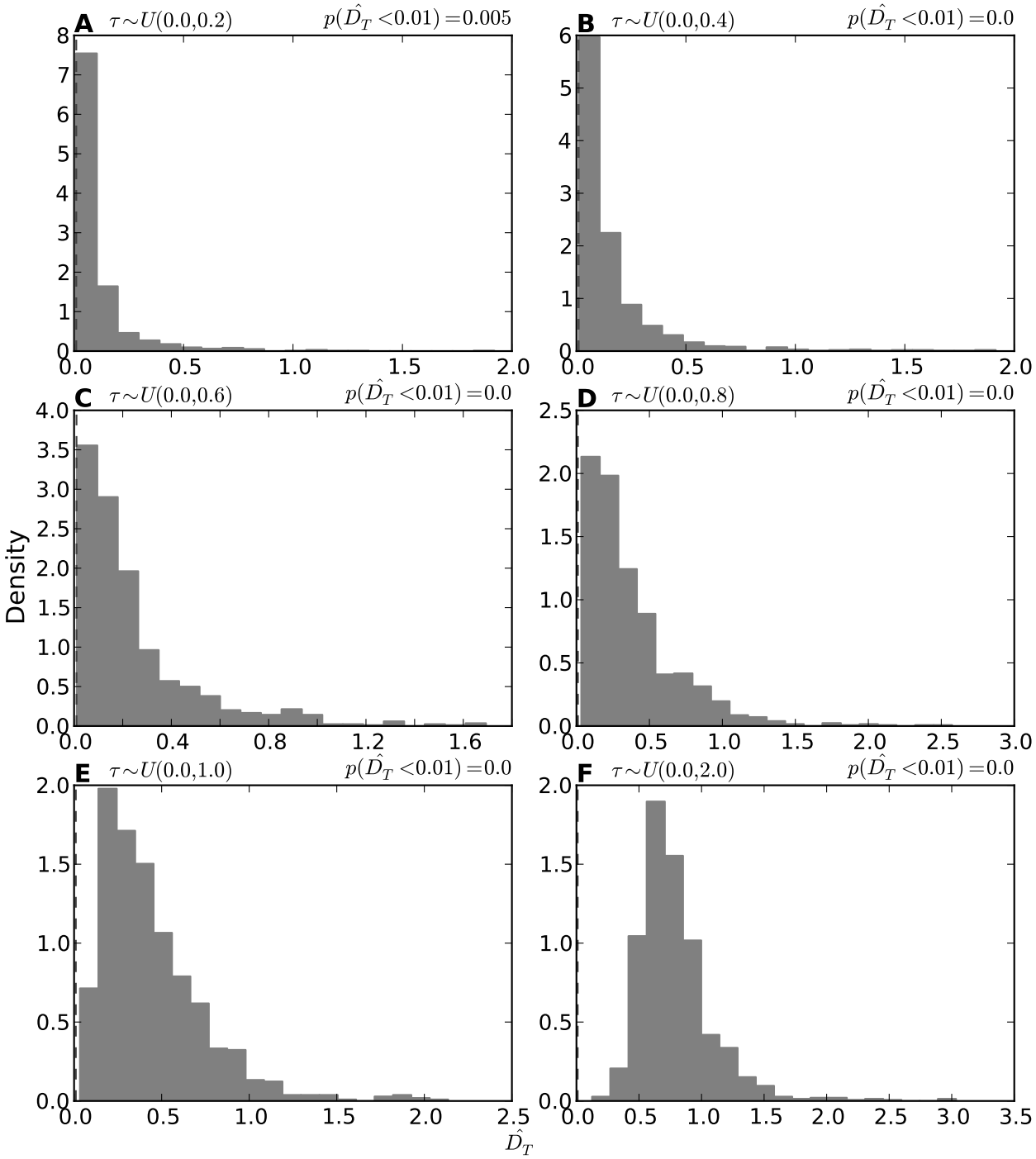

Figure 3.63. The power of model M_{DPP} to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate the estimated dispersion index of divergence times ($\hat{D_T}$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model inferring one divergence event, $p(\hat{D_T} < 0.01)$, given for each data model.

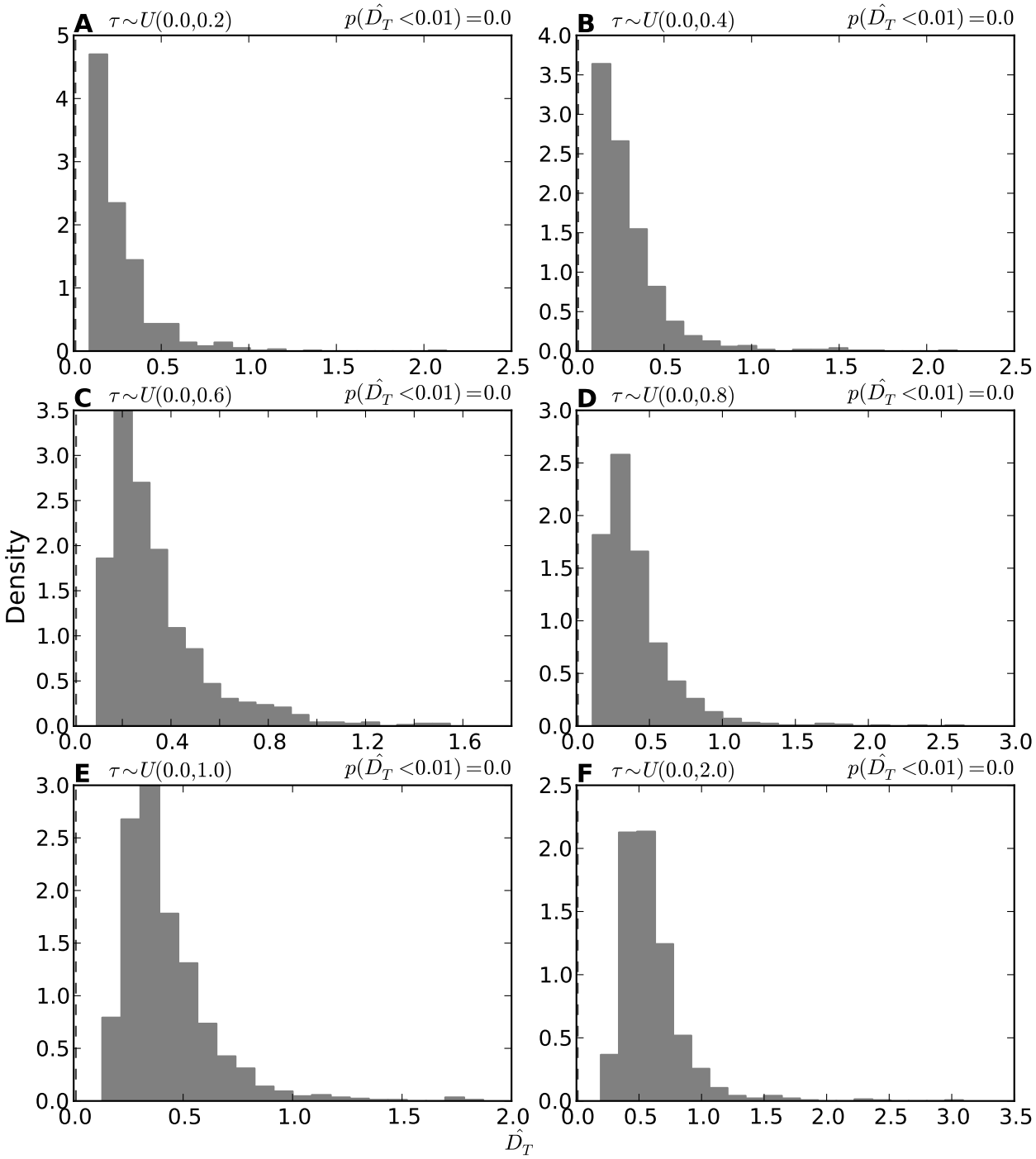

Figure 3.64. The power of model $M_{Uniform}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate the estimated dispersion index of divergence times ($\hat{D_T}$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model inferring one divergence event, $p(\hat{D_T} < 0.01)$, given for each data model.

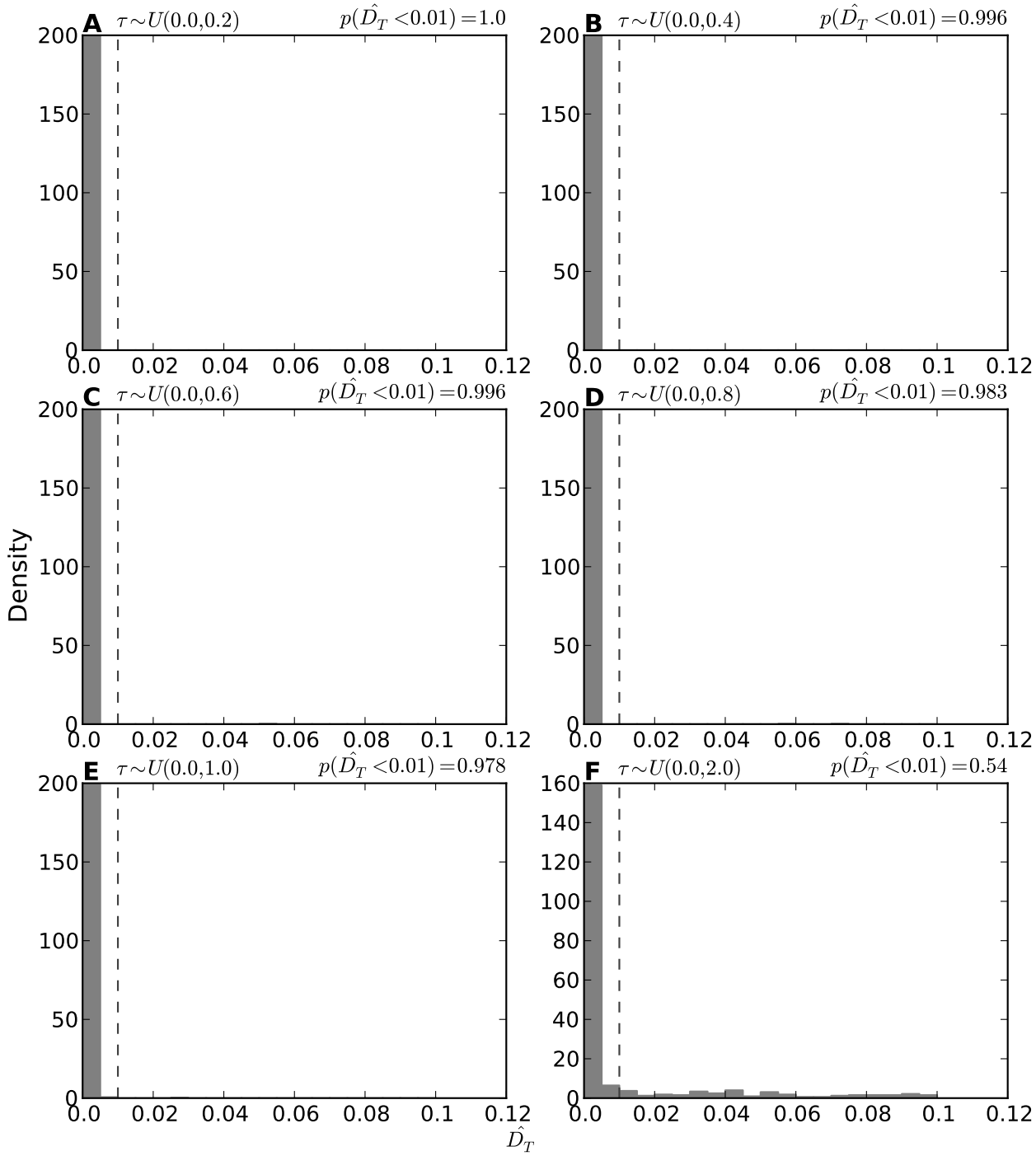

Figure 3.65. The power of model $M_{msBayes}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate the estimated dispersion index of divergence times ($\hat{D}_T$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model inferring one divergence event, $p(\hat{D}_T < 0.01)$, given for each data model.

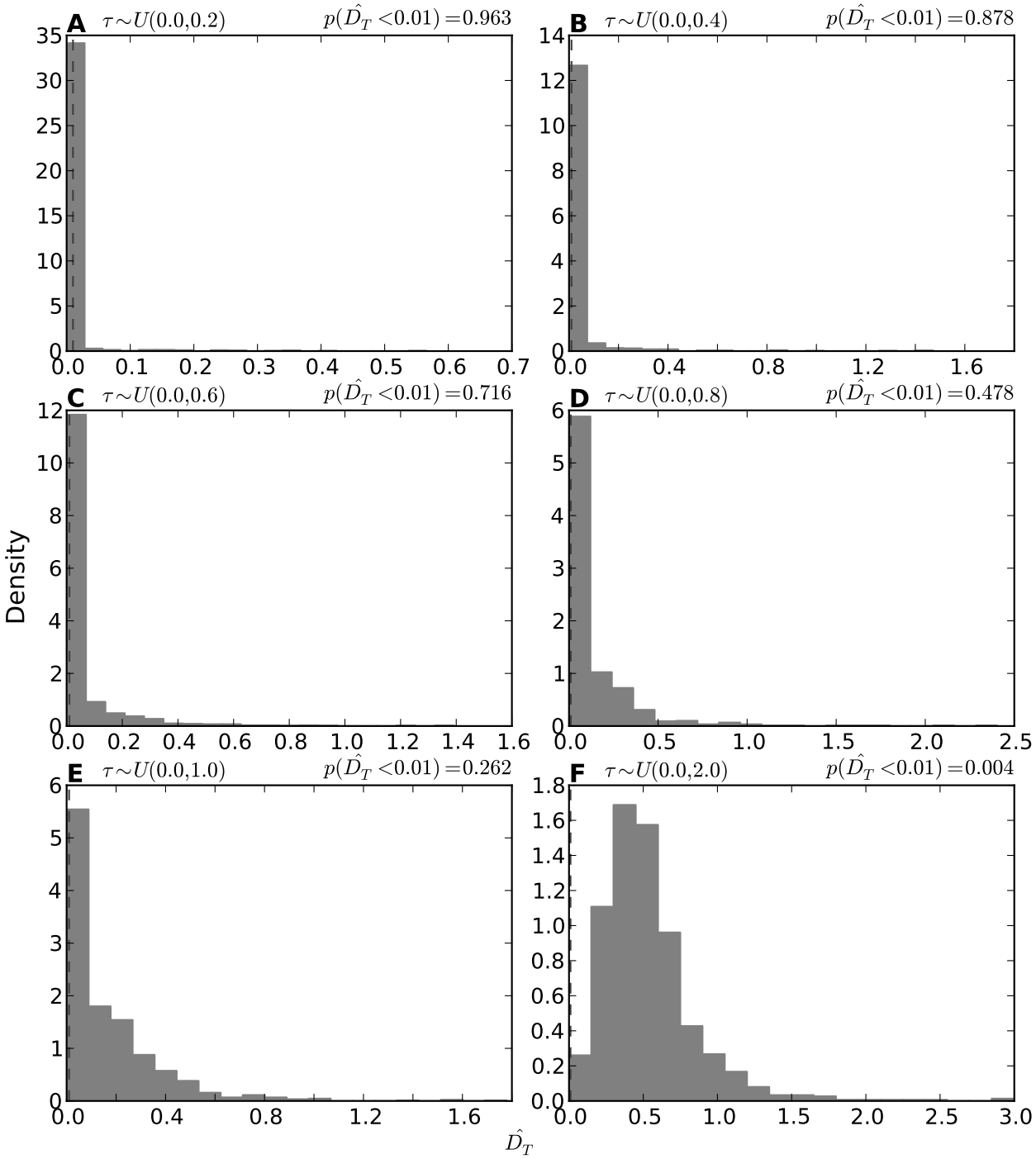

Figure 3.66. The power of model $M_{Ushaped}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate the estimated dispersion index of divergence times ($\hat{D_T}$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model inferring one divergence event, $p(\hat{D_T} < 0.01)$, given for each data model.

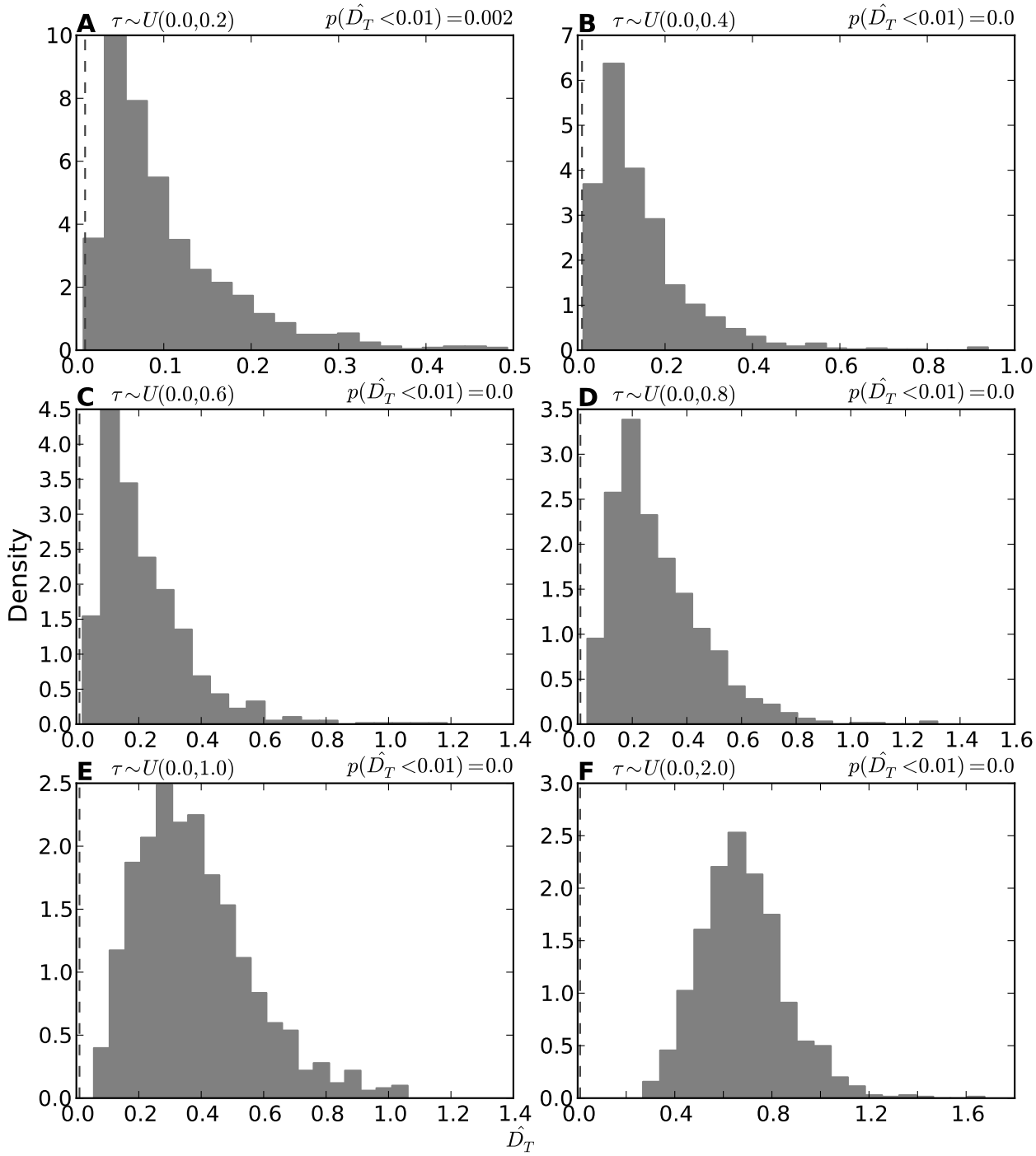

Figure 3.67. The power of model M_{DPP} to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate the estimated dispersion index of divergence times ($\hat{D_T}$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model inferring one divergence event, $p(\hat{D_T} < 0.01)$, given for each data model.

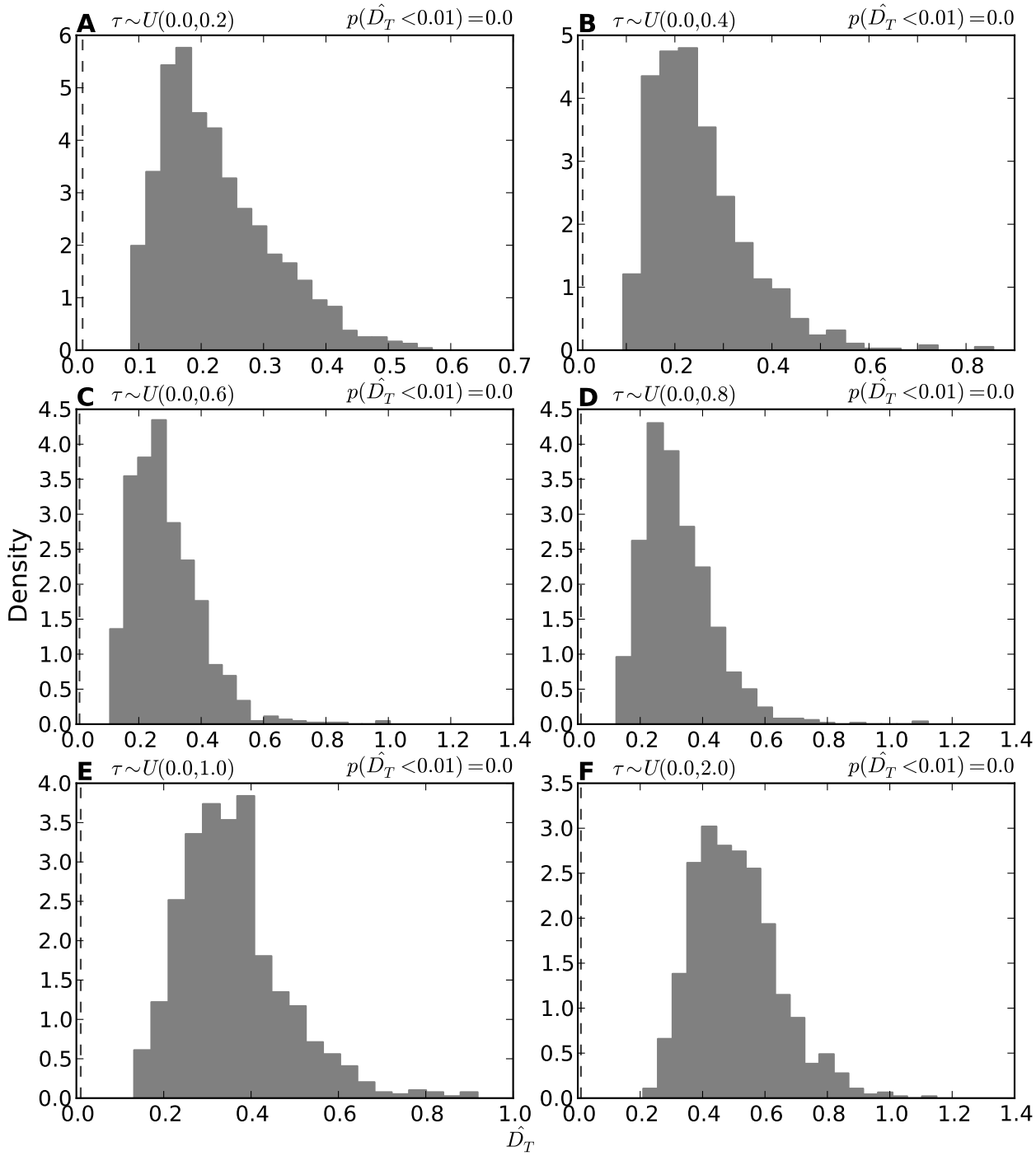

Figure 3.68. The power of model $M_{Uniform}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate the estimated dispersion index of divergence times ($\hat{D}_T$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model inferring one divergence event, $p(\hat{D}_T < 0.01)$, given for each data model.

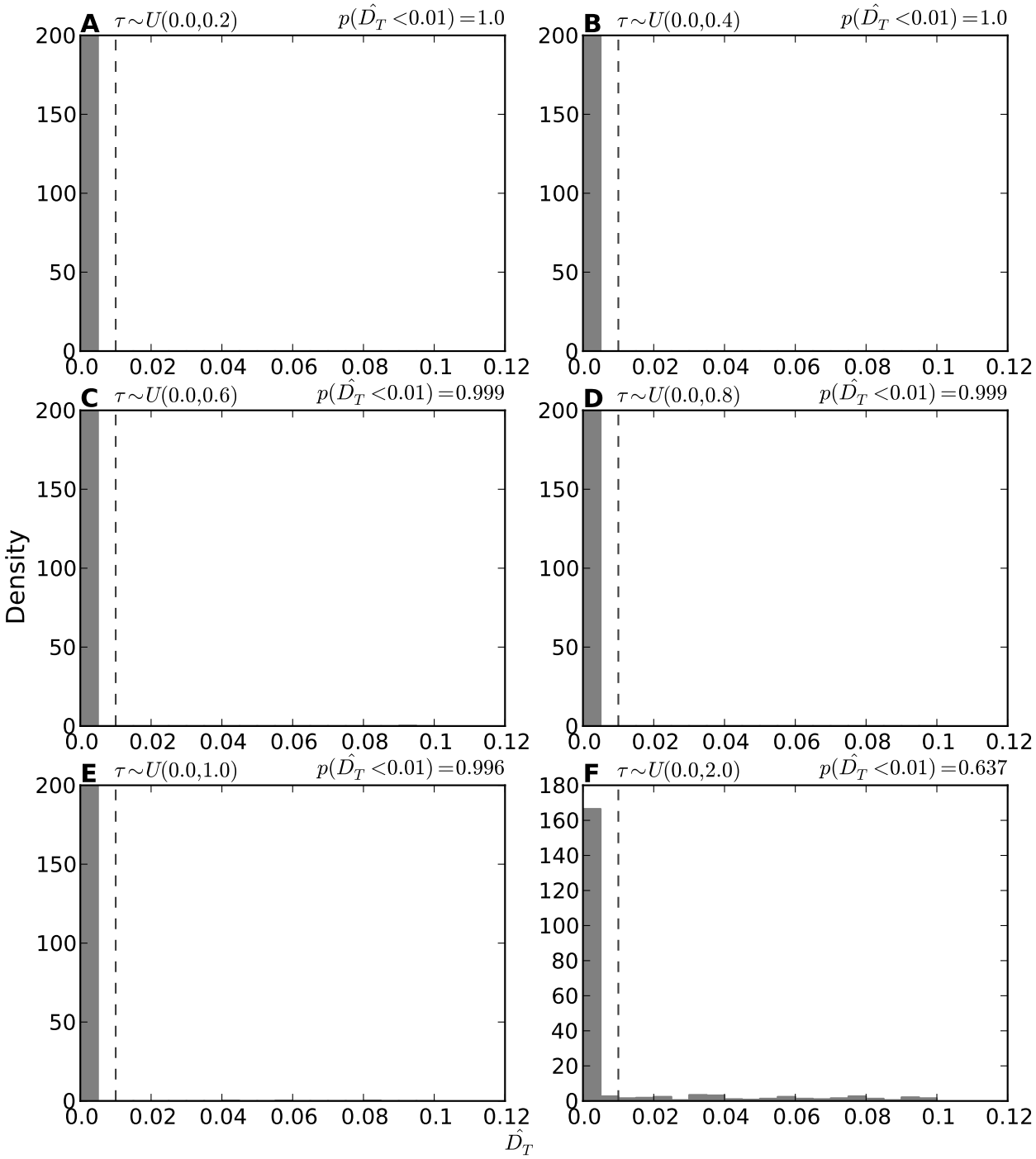

Figure 3.69. The power of model $M_{msBayes}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate the estimated dispersion index of divergence times ($\hat{D}_T$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model inferring one divergence event, $p(\hat{D}_T < 0.01)$, given for each data model.

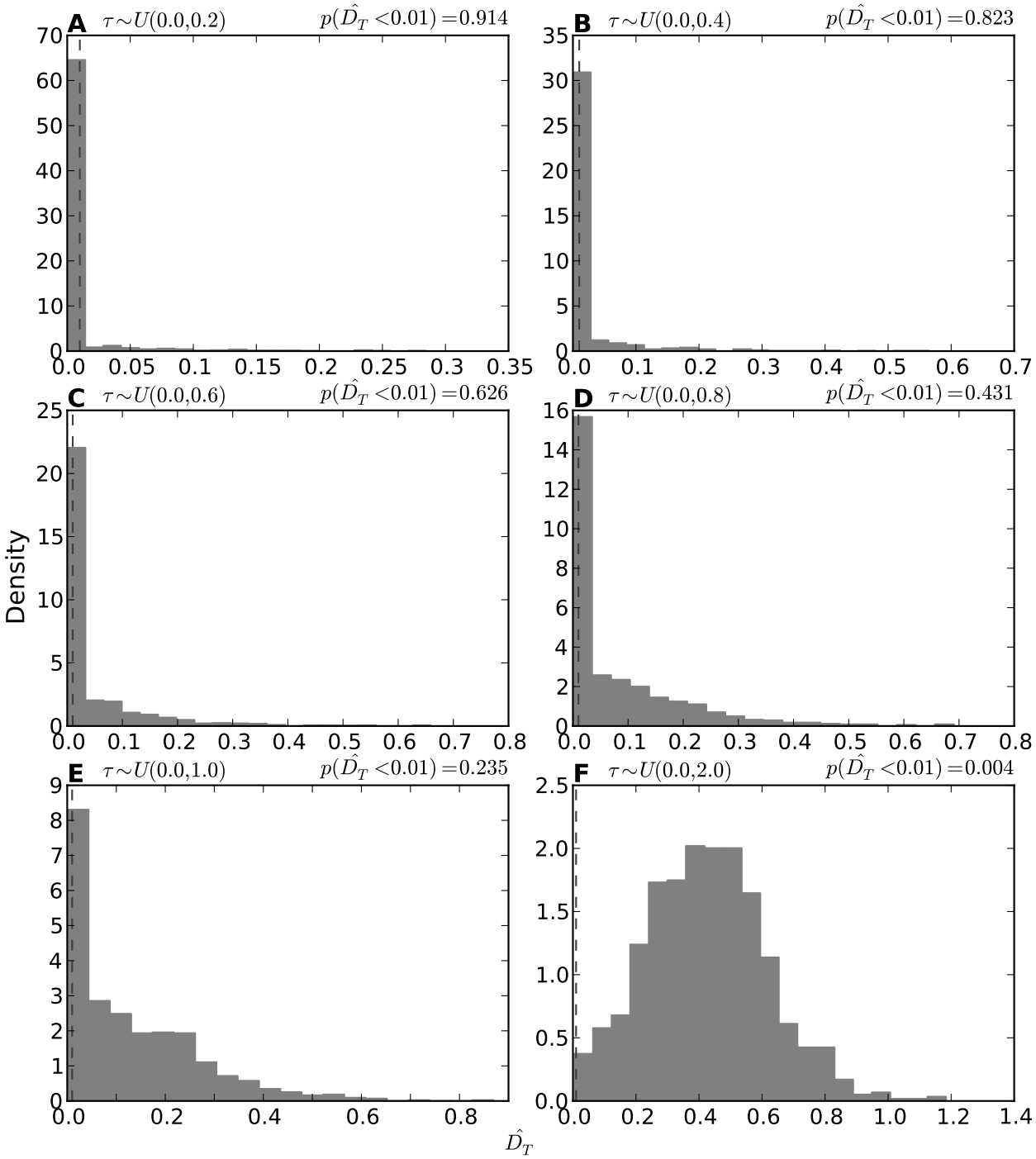

Figure 3.70. The power of model $M_{Ushaped}$ to detect random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate the estimated dispersion index of divergence times ($\hat{D}_T$) from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model inferring one divergence event, $p(\hat{D}_T < 0.01)$, given for each data model.

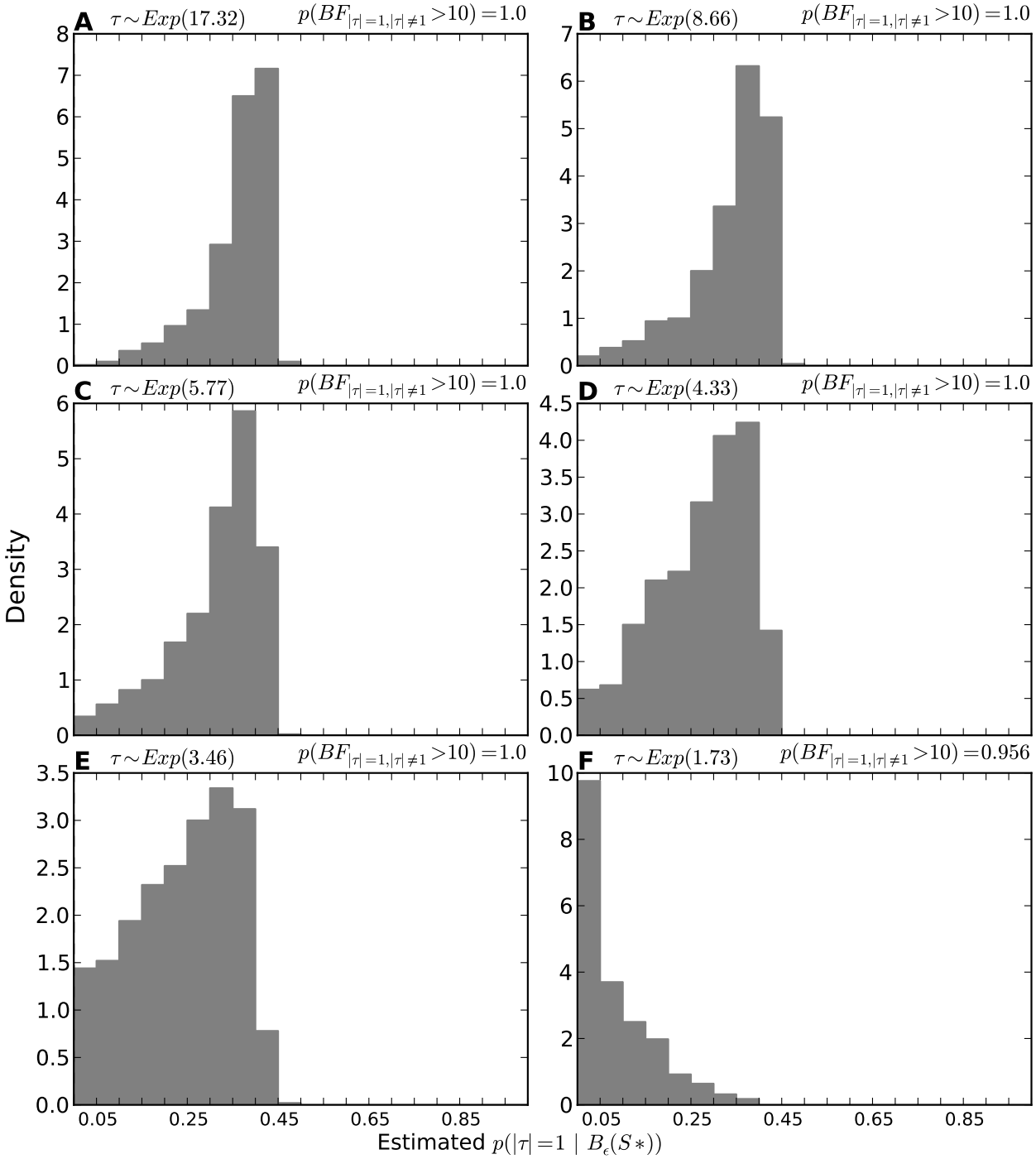

Figure 3.71. The tendency of model M_{DPP} to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(|\boldsymbol{\tau}| = 1|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{|\boldsymbol{\tau}|=1,|\boldsymbol{\tau}|\neq 1} > 10)$, given for each data model.

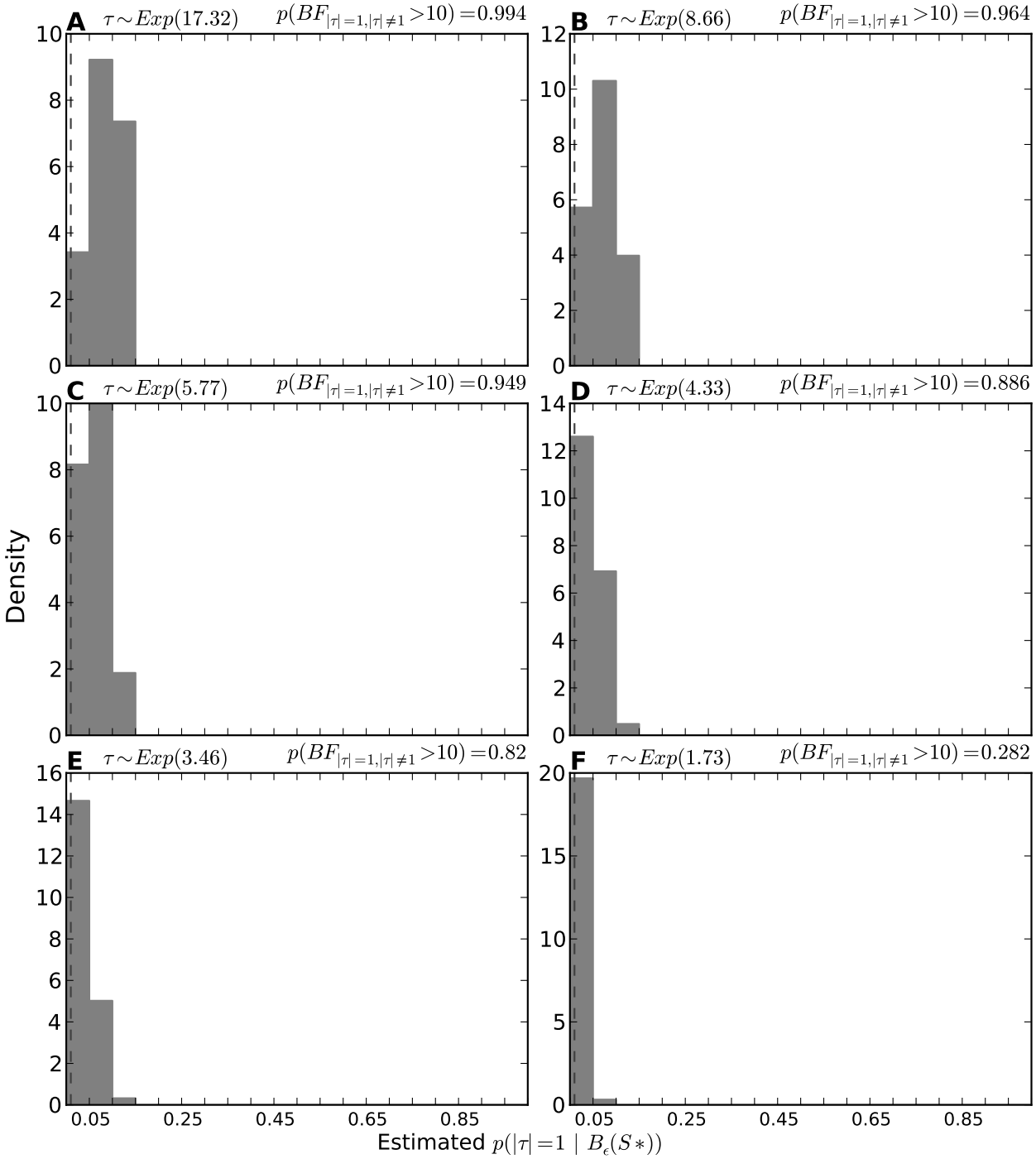

Figure 3.72. The tendency of model $M_{Uniform}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(|\boldsymbol{\tau}| = 1|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{|\boldsymbol{\tau}|=1,|\boldsymbol{\tau}|\neq 1} > 10)$, given for each data model.

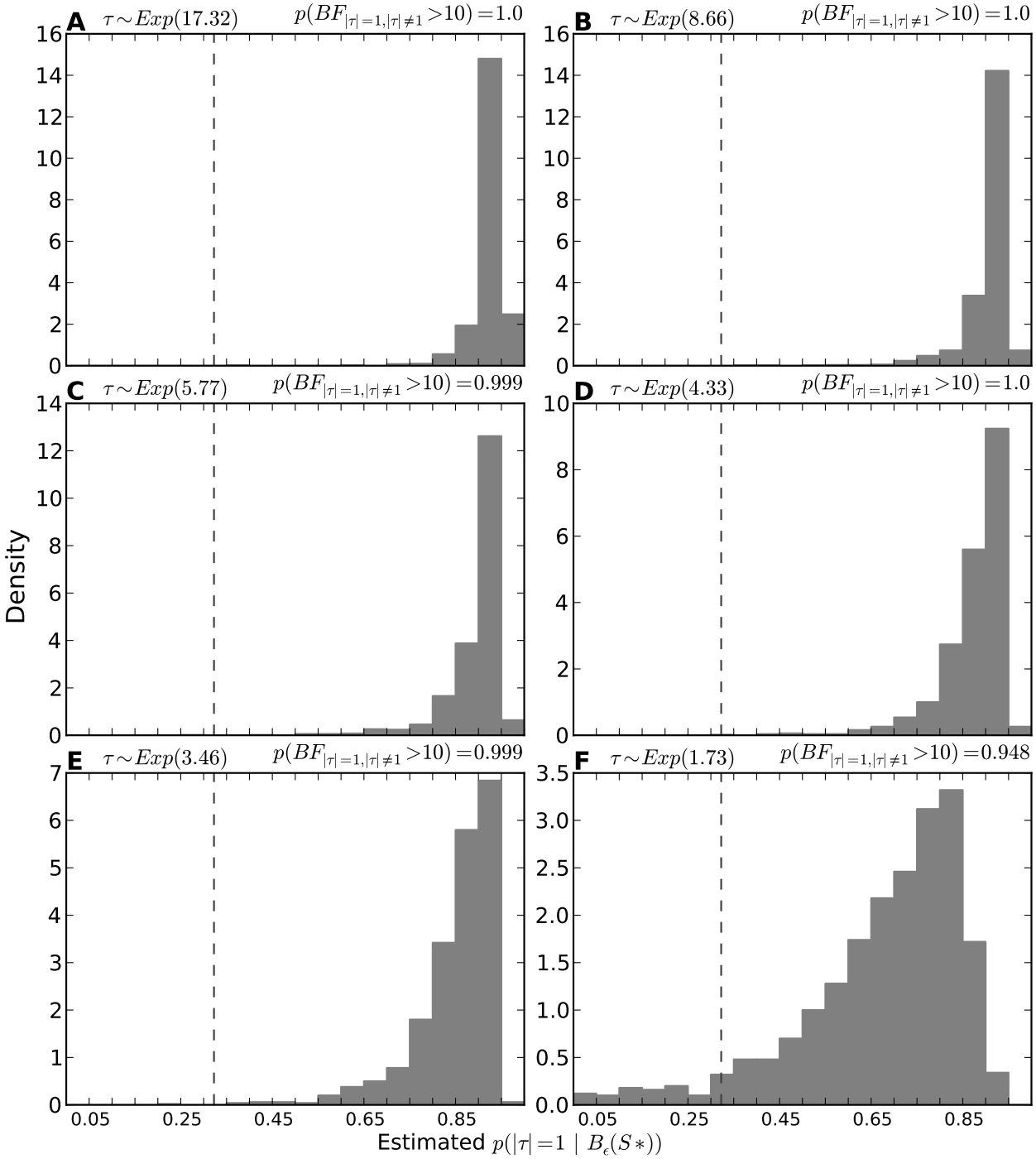

Figure 3.73. The tendency of model $M_{msBayes}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(|\boldsymbol{\tau}| = 1|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{|\boldsymbol{\tau}|=1,|\boldsymbol{\tau}|\neq 1} > 10)$, given for each data model.

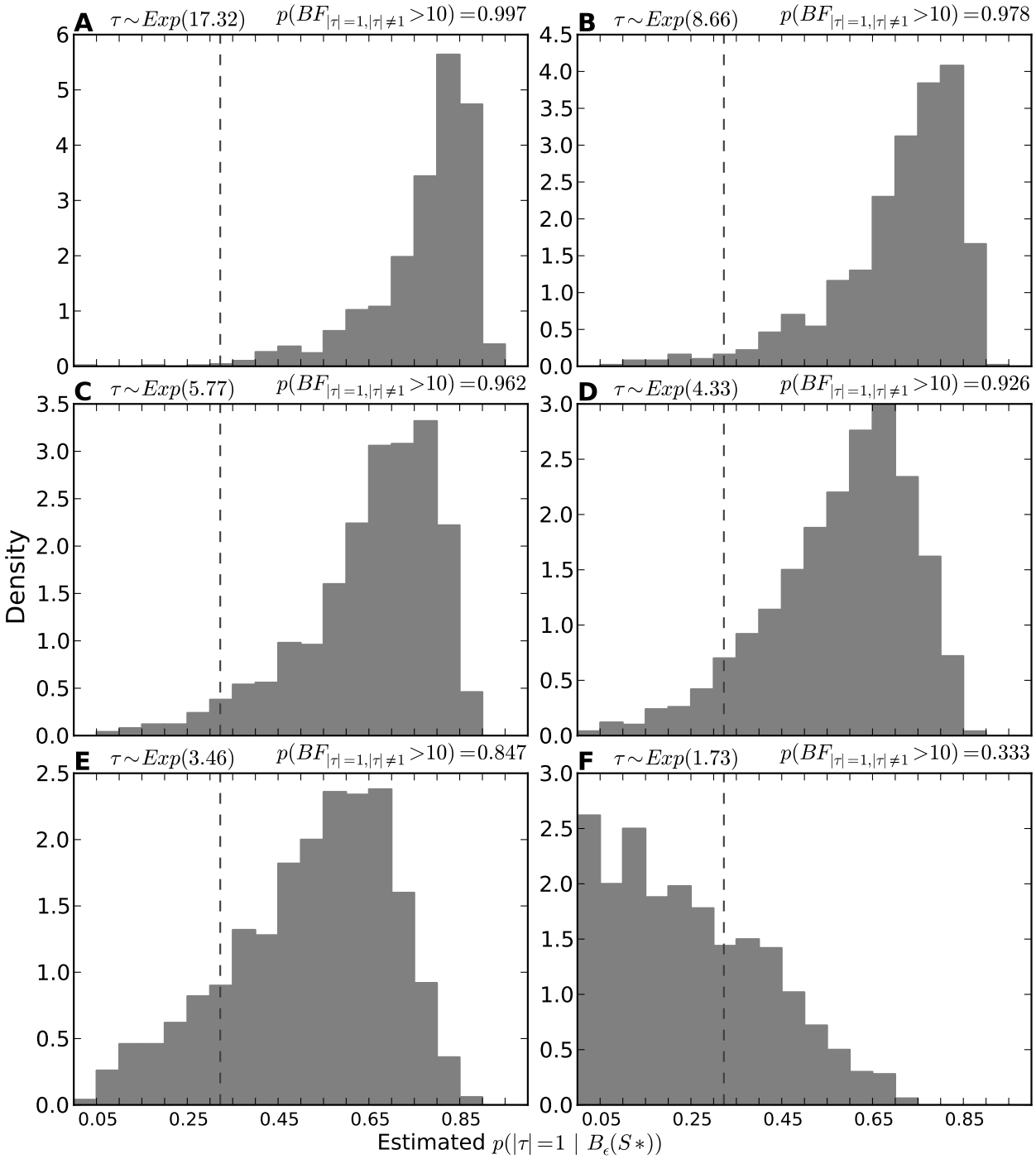

Figure 3.74. The tendency of model $M_{Ushaped}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(|\boldsymbol{\tau}| = 1|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{|\boldsymbol{\tau}|=1,|\boldsymbol{\tau}|\neq 1} > 10)$, given for each data model.

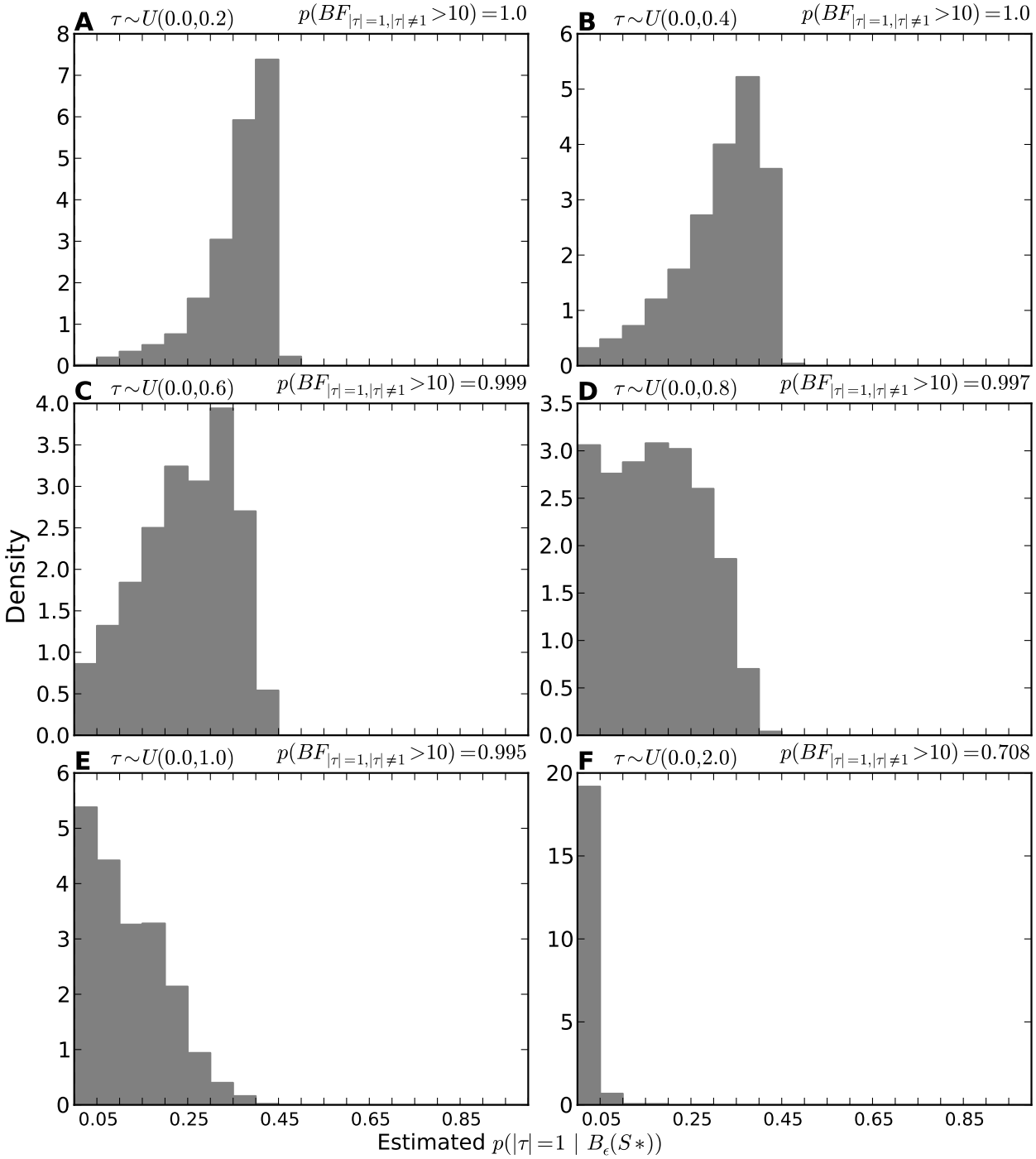

Figure 3.75. The tendency of model M_{DPP} to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(|\boldsymbol{\tau}| = 1|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{|\boldsymbol{\tau}|=1,|\boldsymbol{\tau}|\neq 1} > 10)$, given for each data model.

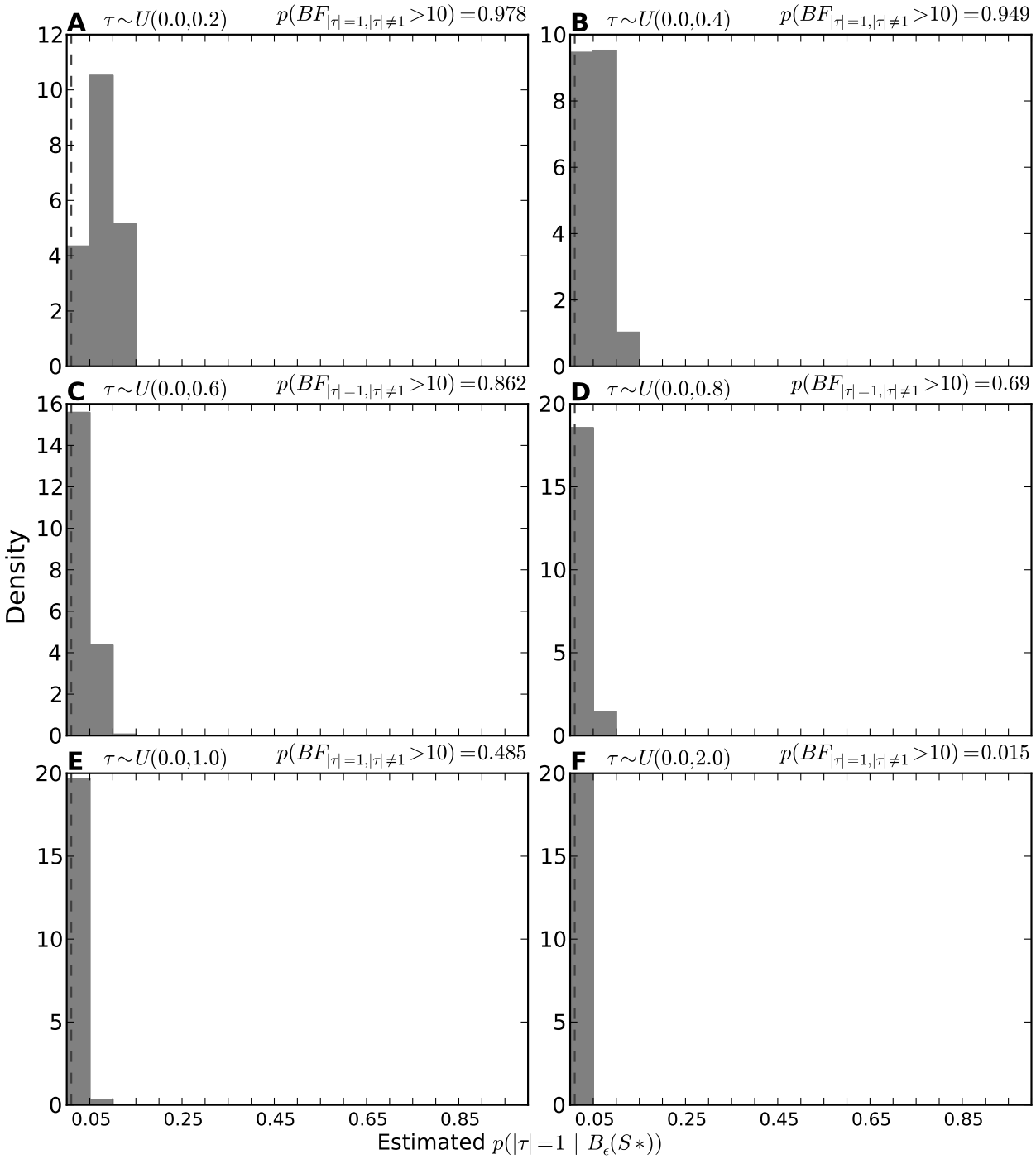

Figure 3.76. The tendency of model $M_{Uniform}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(|\boldsymbol{\tau}| = 1|B_{\epsilon}(\mathbf{S}^{*}))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{|\boldsymbol{\tau}|=1,|\boldsymbol{\tau}|\neq 1} > 10)$, given for each data model.

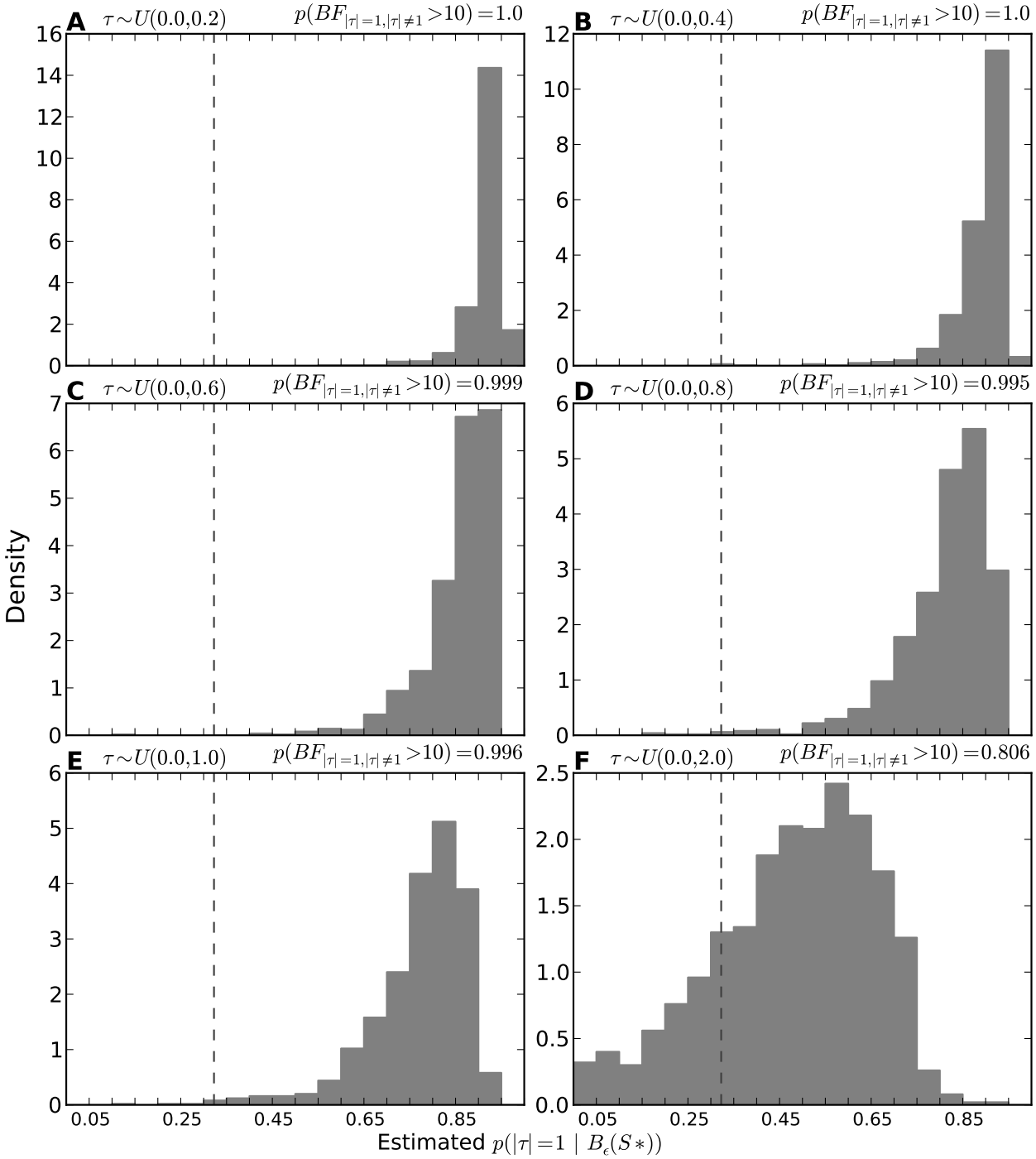

Figure 3.77. The tendency of model $M_{msBayes}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(|\boldsymbol{\tau}| = 1|B_{\epsilon}(\mathbf{S}^{*}))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{|\boldsymbol{\tau}|=1,|\boldsymbol{\tau}|\neq 1} > 10)$, given for each data model.

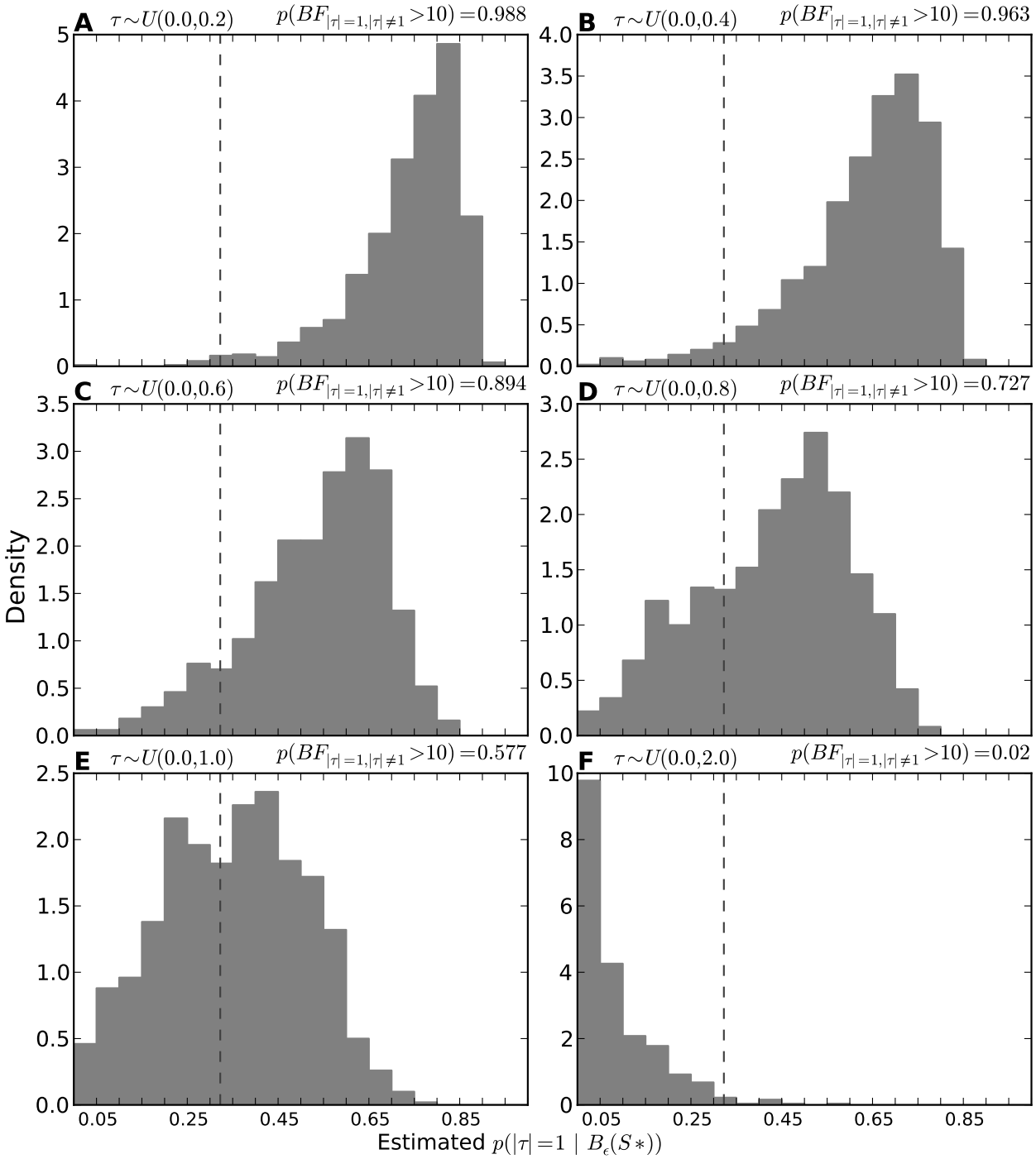

Figure 3.78. The tendency of model $M_{Ushaped}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(|\boldsymbol{\tau}| = 1|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{|\boldsymbol{\tau}|=1,|\boldsymbol{\tau}|\neq 1} > 10)$, given for each data model.

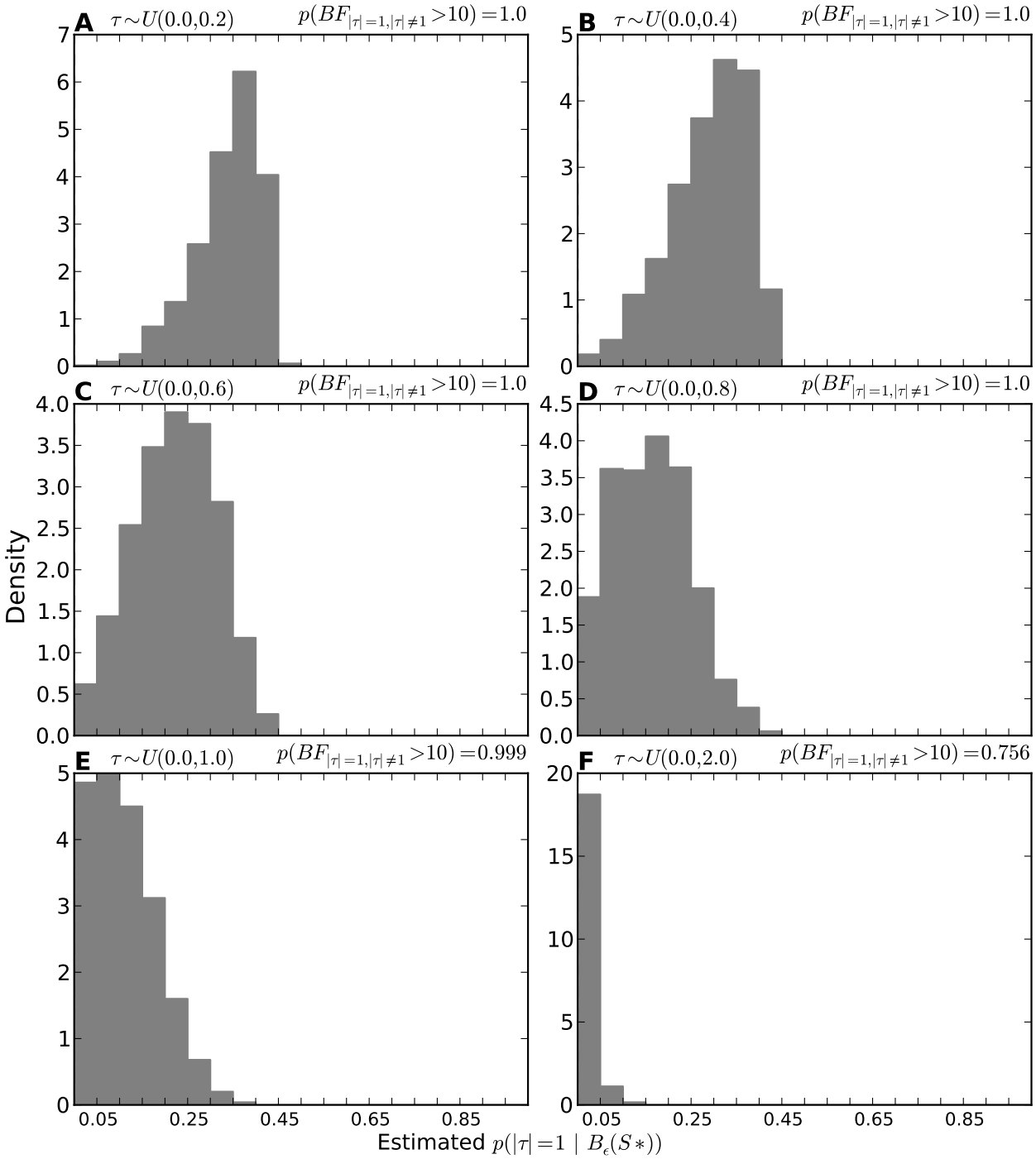

Figure 3.79. The tendency of model M_{DPP} to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(|\boldsymbol{\tau}| = 1|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{|\boldsymbol{\tau}|=1,|\boldsymbol{\tau}|\neq 1} > 10)$, given for each data model.

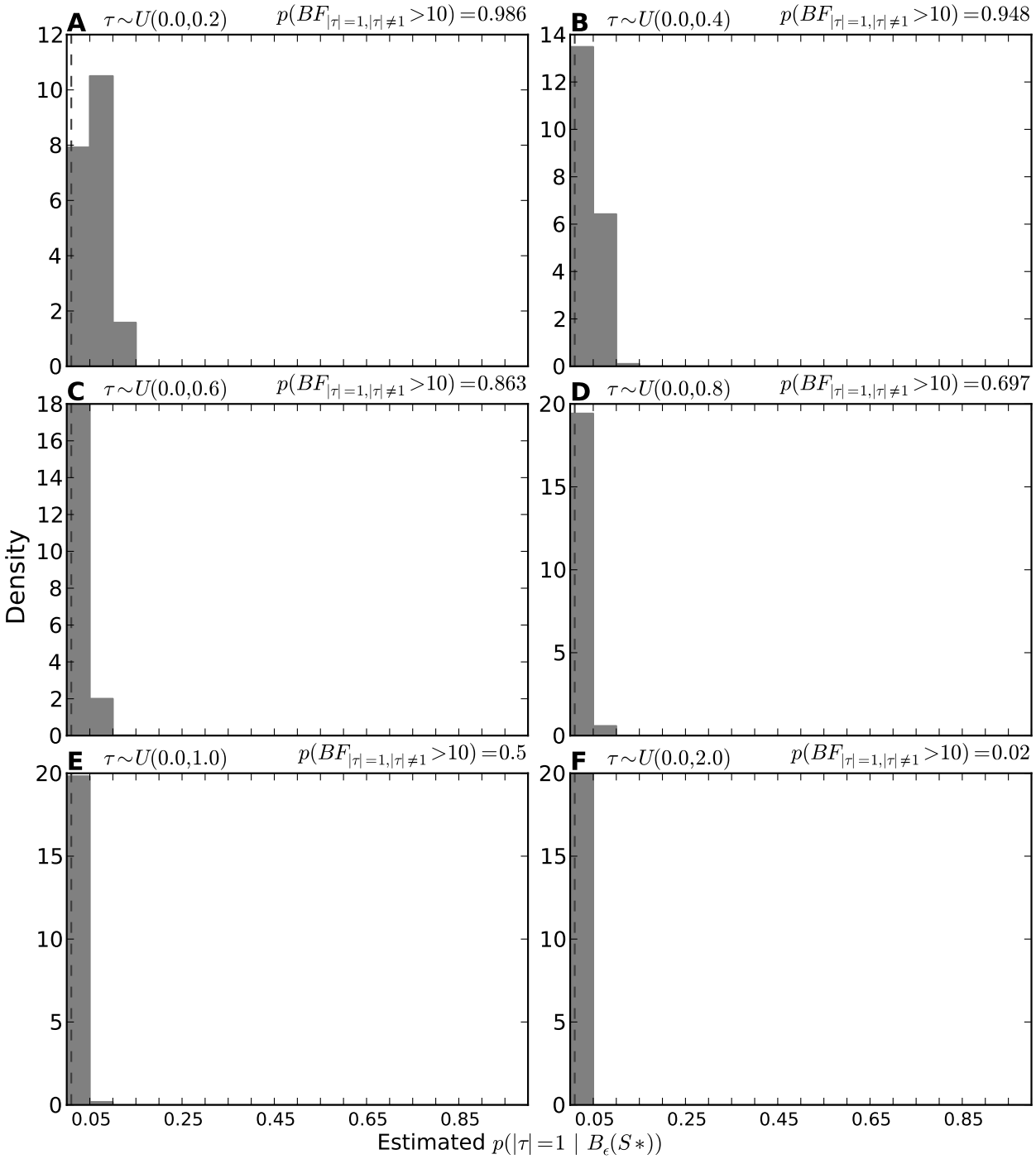

Figure 3.80. The tendency of model $M_{Uniform}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(|\boldsymbol{\tau}| = 1|B_{\epsilon}(\mathbf{S}^{*}))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{|\boldsymbol{\tau}|=1,|\boldsymbol{\tau}|\neq 1} > 10)$, given for each data model.

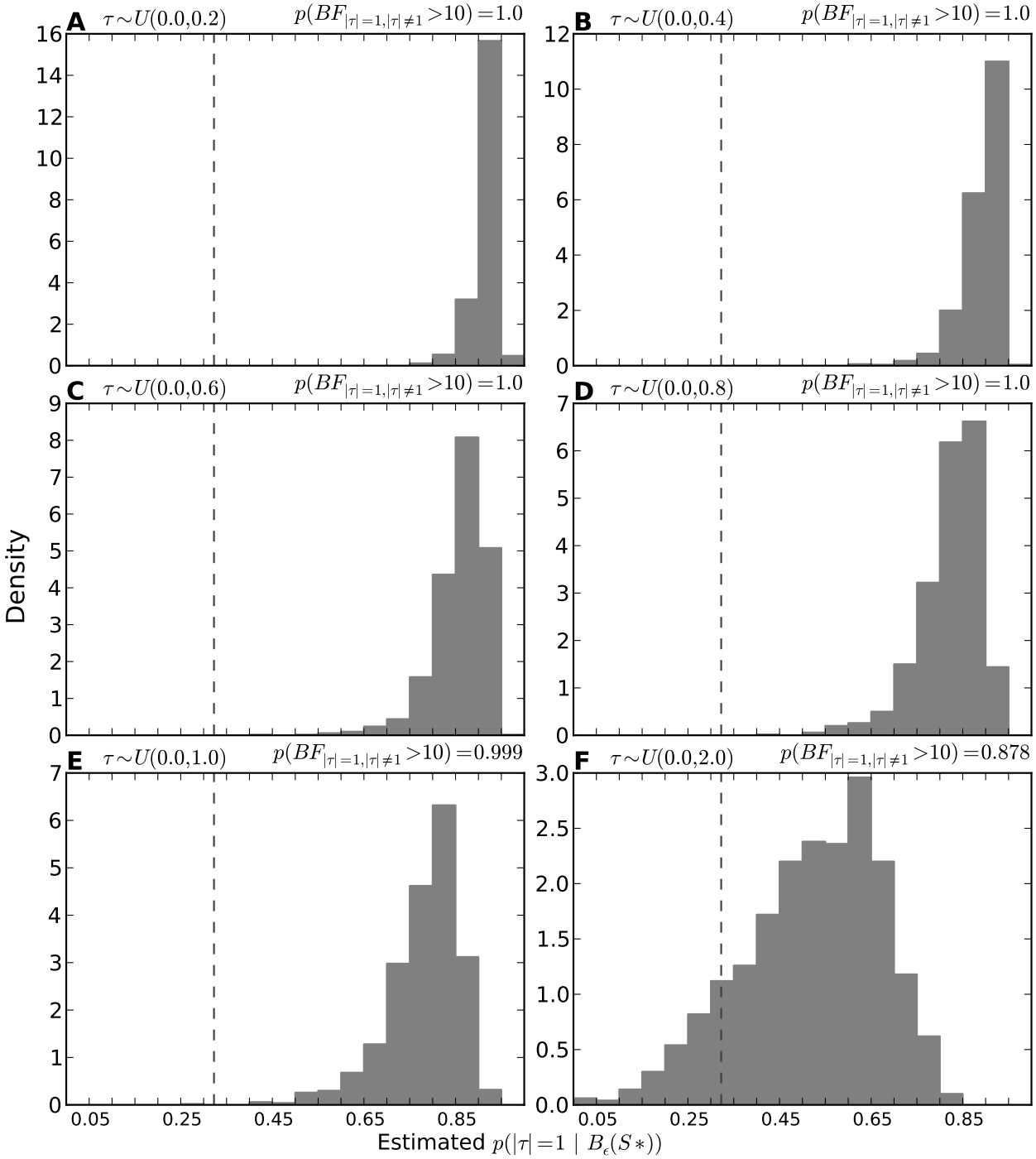

Figure 3.81. The tendency of model $M_{msBayes}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(|\boldsymbol{\tau}| = 1|B_{\epsilon}(\mathbf{S}^{*}))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{|\boldsymbol{\tau}|=1,|\boldsymbol{\tau}|\neq 1} > 10)$, given for each data model.

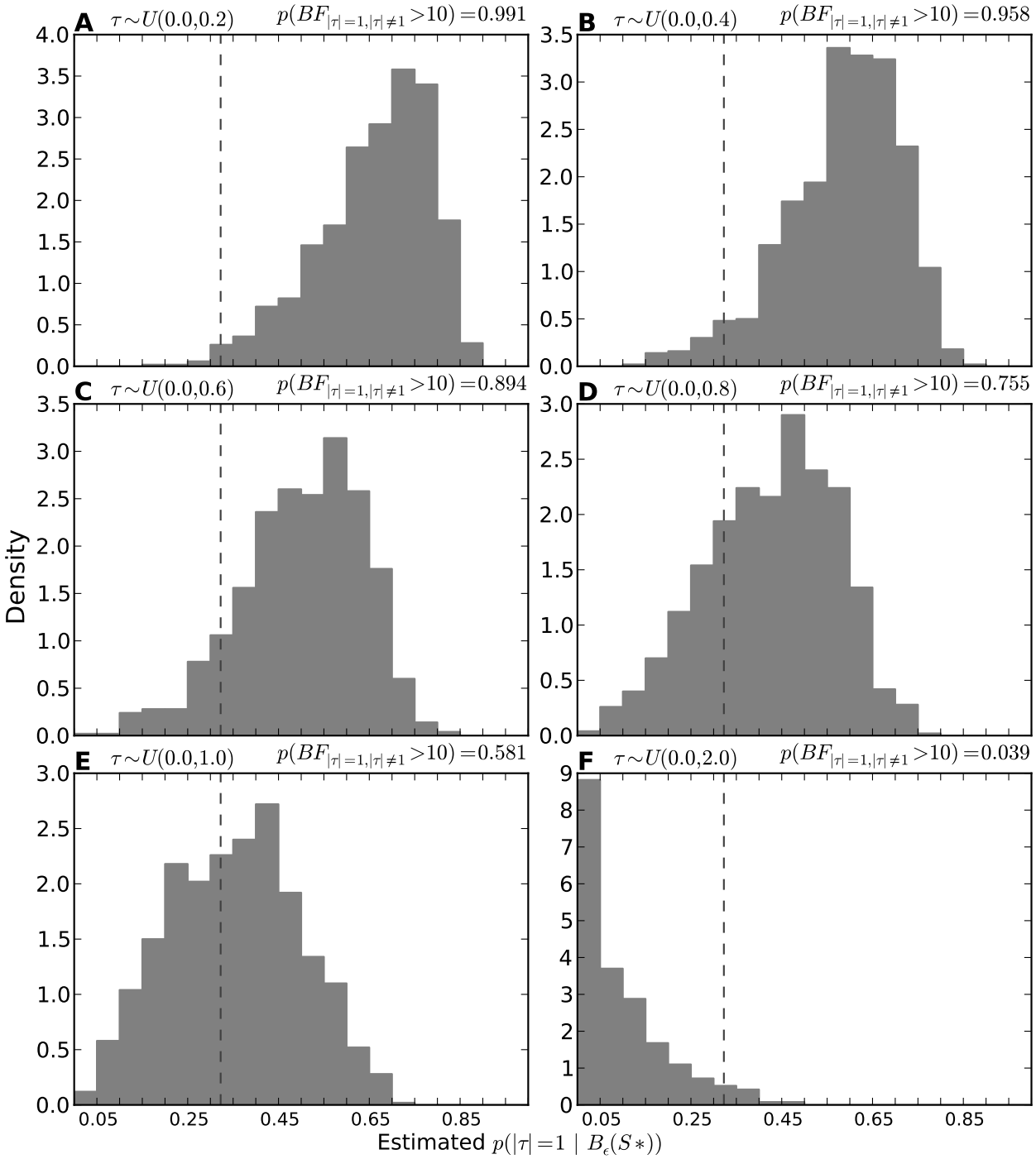

Figure 3.82. The tendency of model $M_{Ushaped}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(|\boldsymbol{\tau}| = 1|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{|\boldsymbol{\tau}|=1,|\boldsymbol{\tau}|\neq 1} > 10)$, given for each data model.

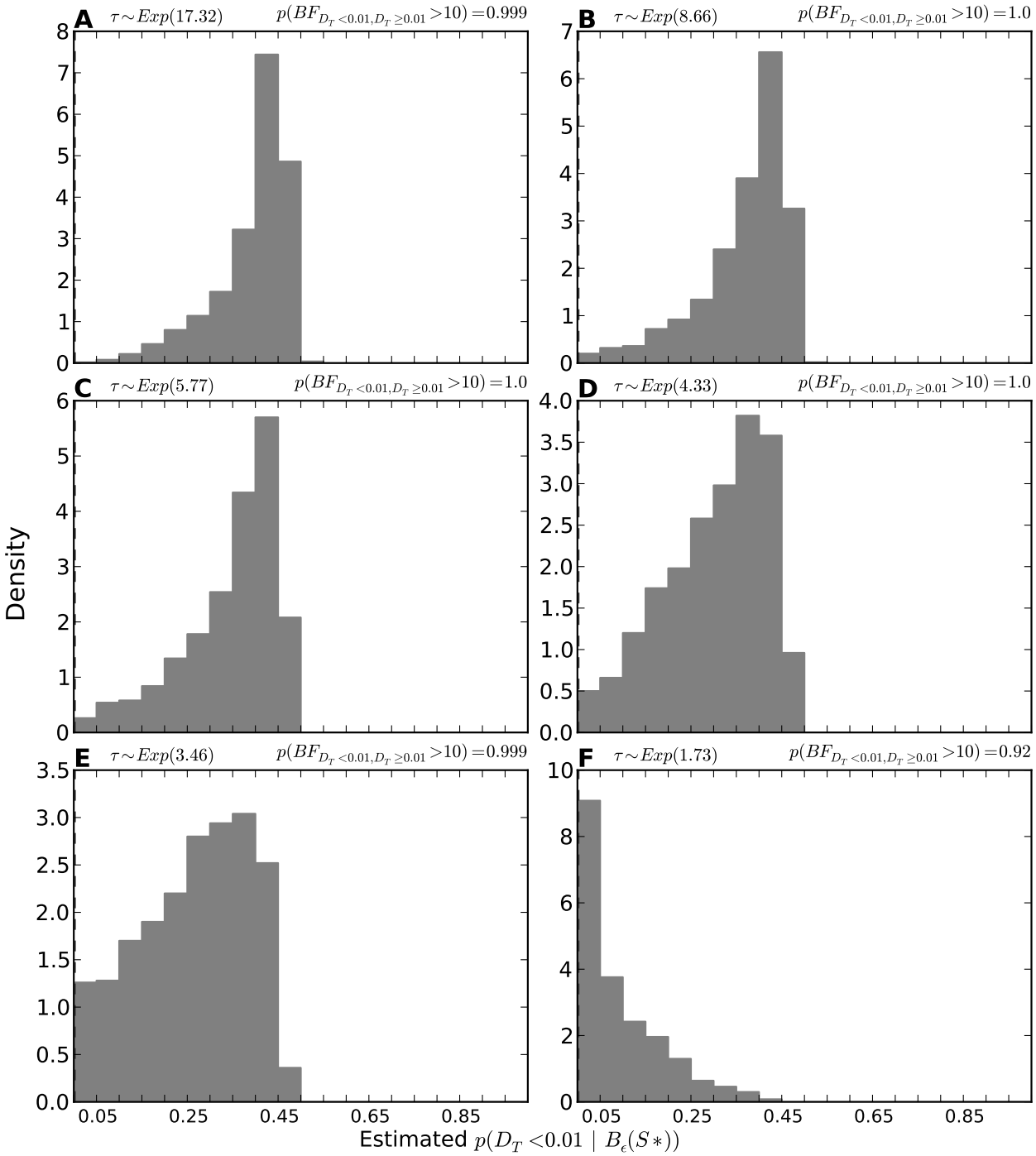

Figure 3.83. The tendency of model M_{DPP} to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{D_T<0.01, D_T \geq 0.01} > 10)$, given for each data model.

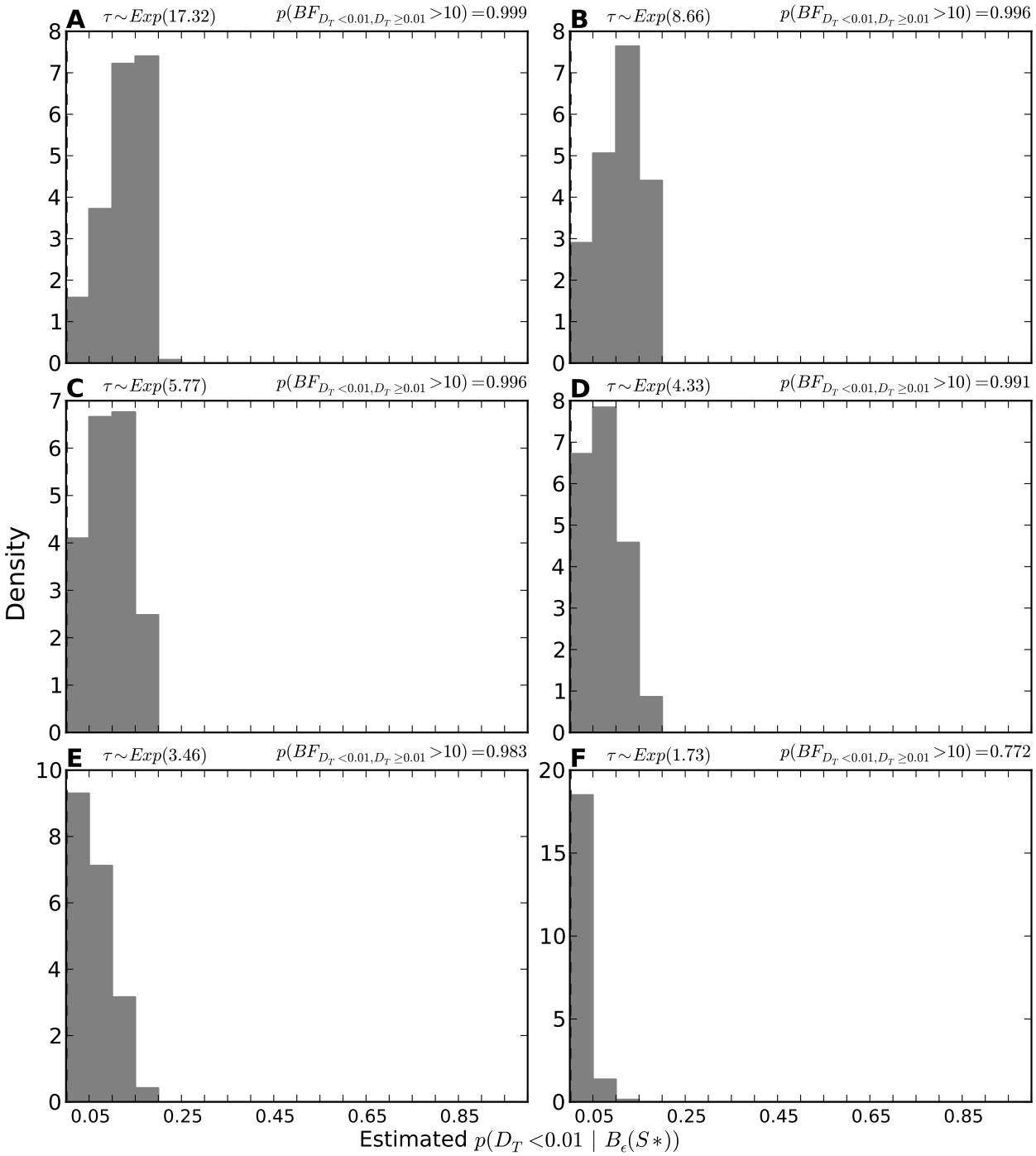

Figure 3.84. The tendency of model $M_{Uniform}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{D_T<0.01, D_T \geq 0.01} > 10)$, given for each data model.

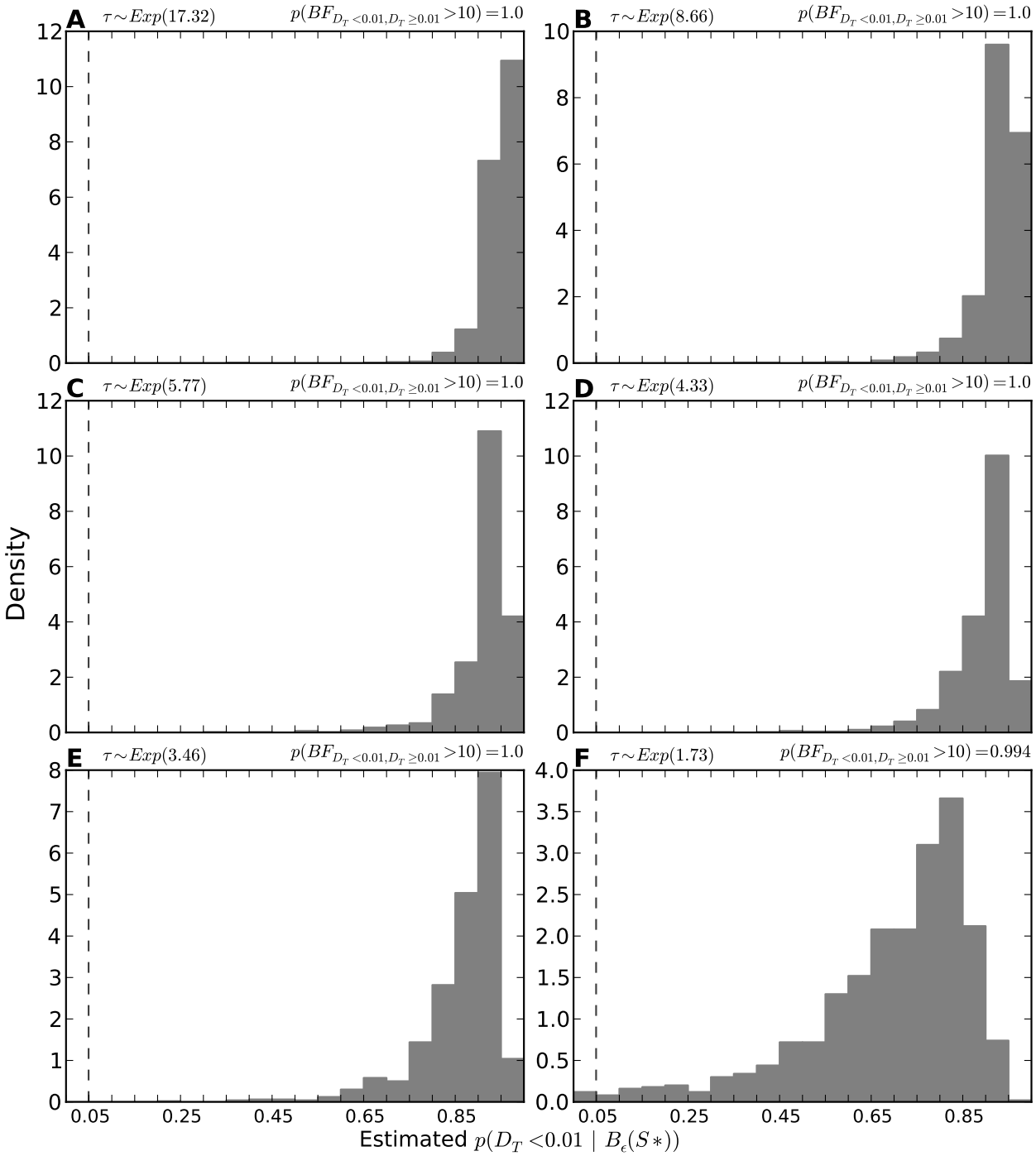

Figure 3.85. The tendency of model $M_{msBayes}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{D_T<0.01, D_T \geq 0.01} > 10)$, given for each data model.

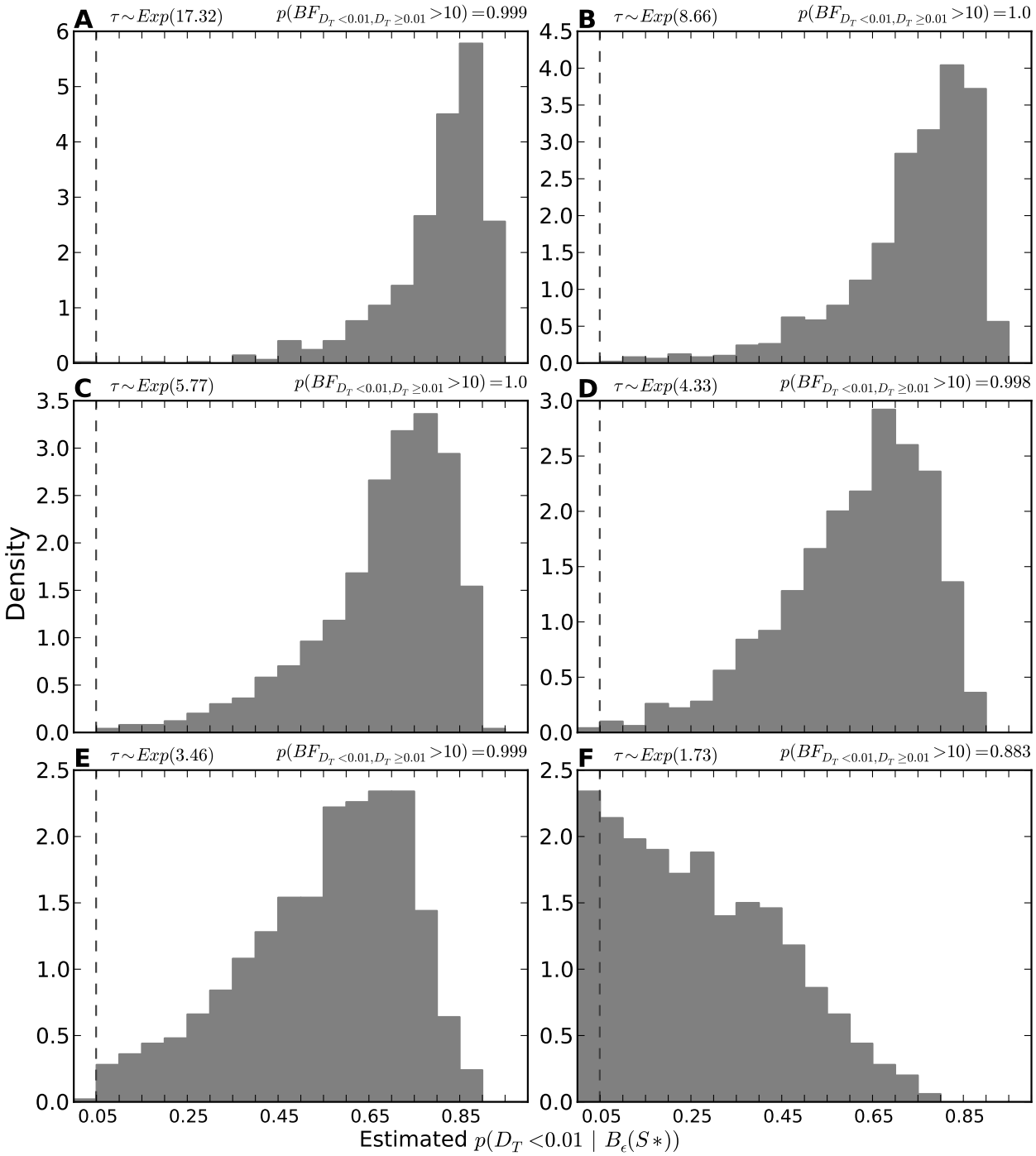

Figure 3.86. The tendency of model $M_{Ushaped}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Exp}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Exp}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{D_T<0.01, D_T \geq 0.01} > 10)$, given for each data model.

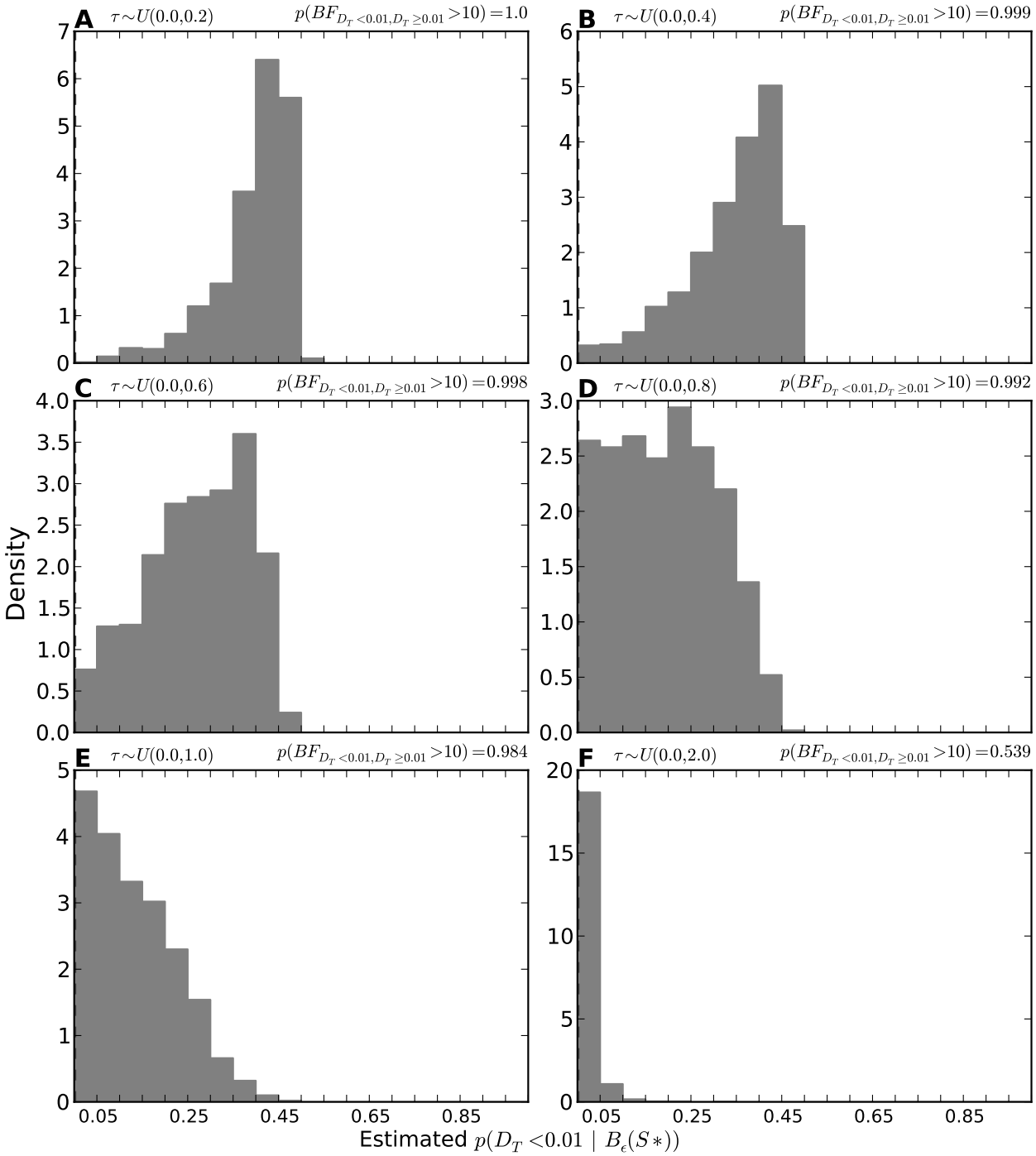

Figure 3.87. The tendency of model M_{DPP} to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{D_T<0.01,D_T\geq0.01} > 10)$, given for each data model.

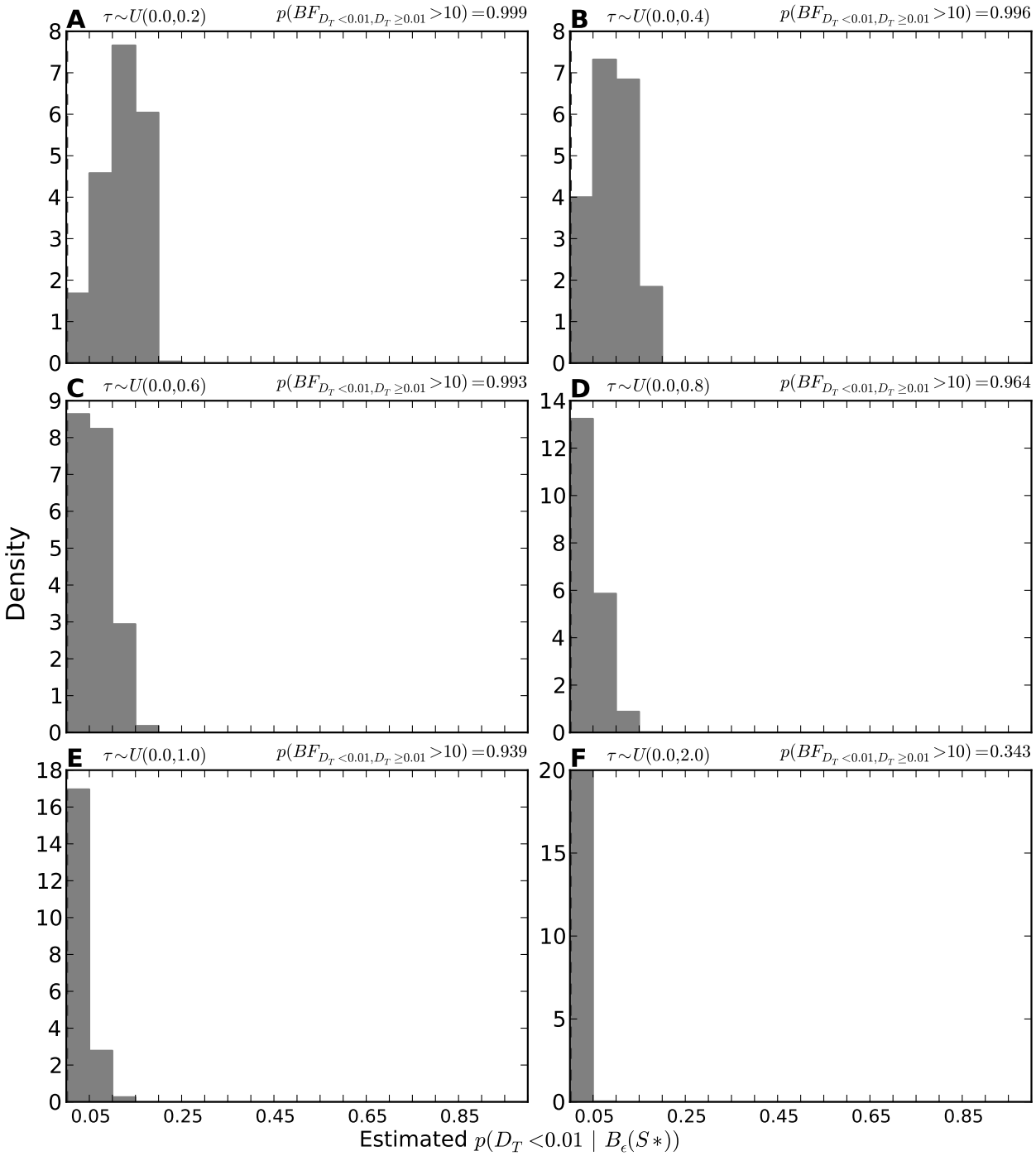

Figure 3.88. The tendency of model $M_{Uniform}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{D_T<0.01,D_T\geq0.01} > 10)$, given for each data model.

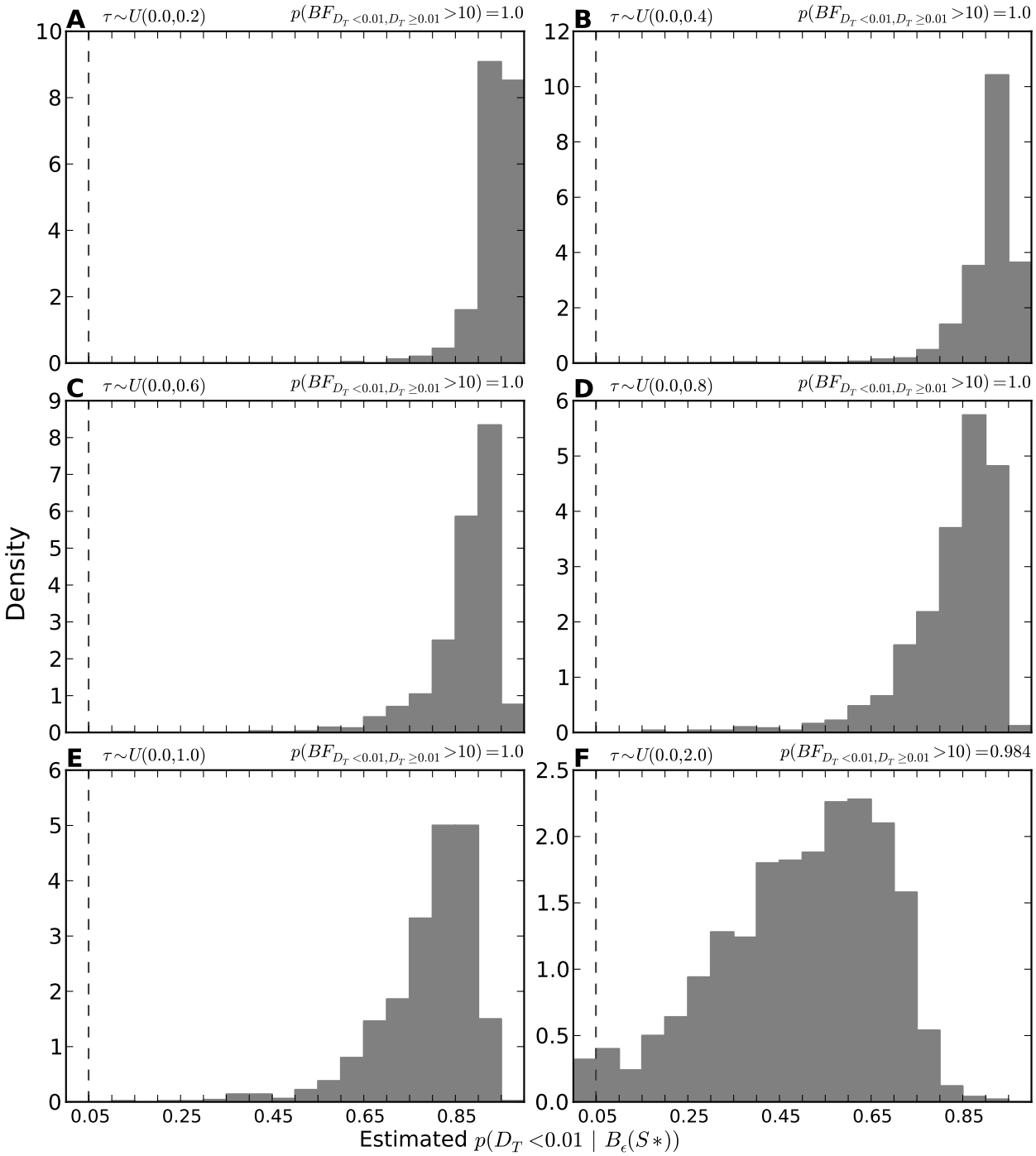

Figure 3.89. The tendency of model $M_{msBayes}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{D_T<0.01, D_T \geq 0.01} > 10)$, given for each data model.

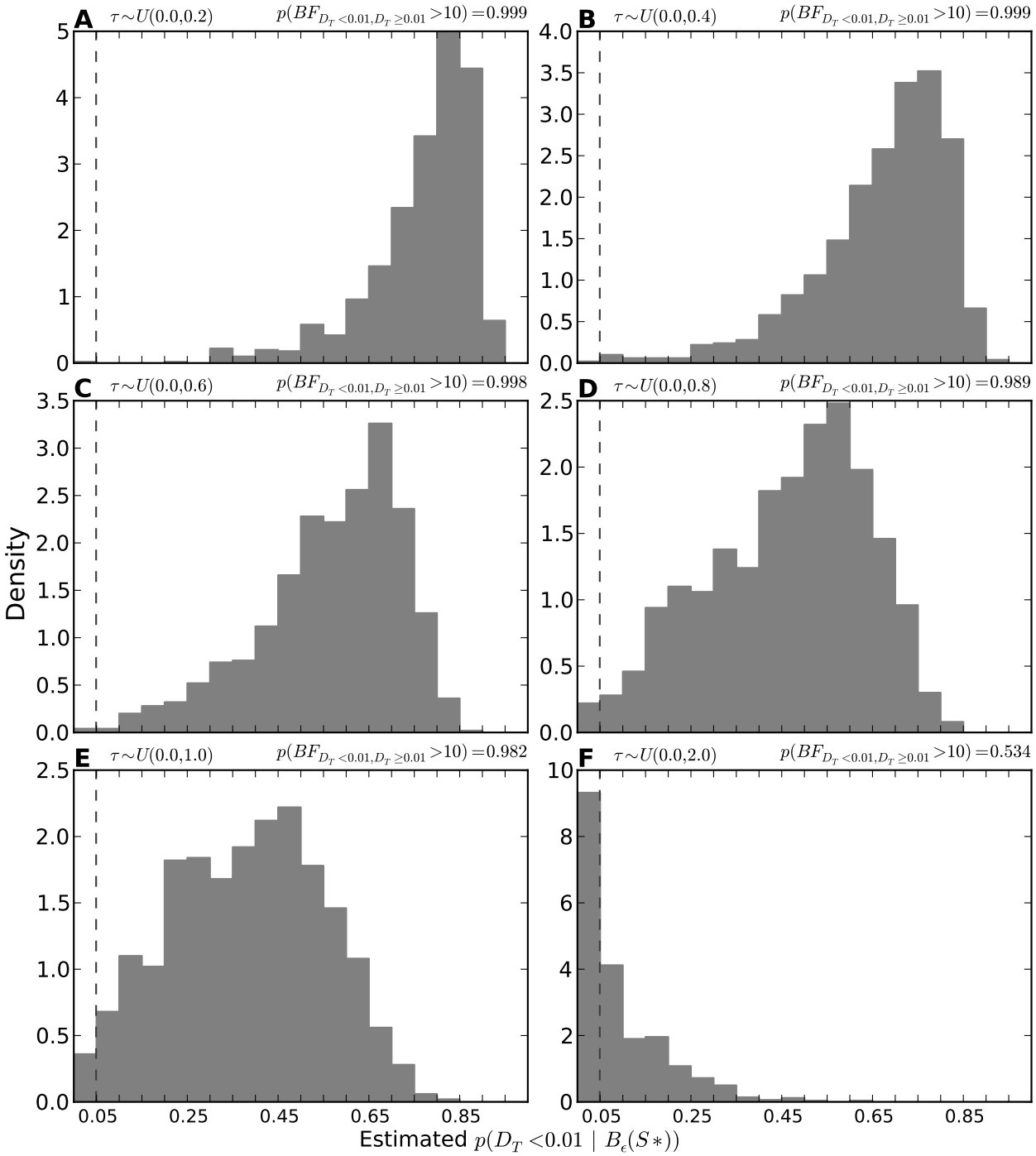

Figure 3.90. The tendency of model $M_{Ushaped}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{Uniform}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{Uniform}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{D_T<0.01,D_T\geq 0.01} > 10)$, given for each data model.

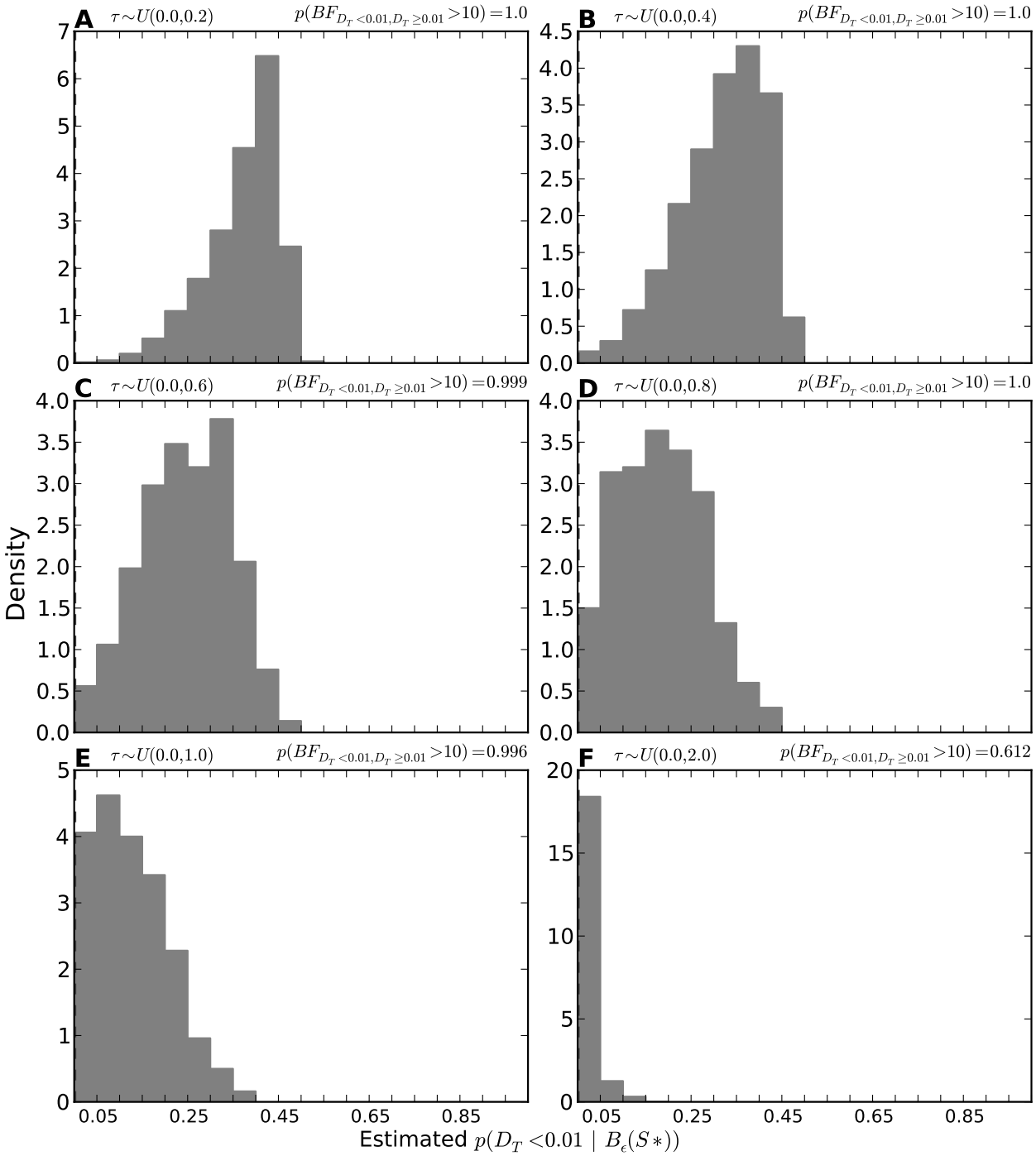

Figure 3.91. The tendency of model M_{DPP} to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{D_T<0.01, D_T \geq 0.01} > 10)$, given for each data model.

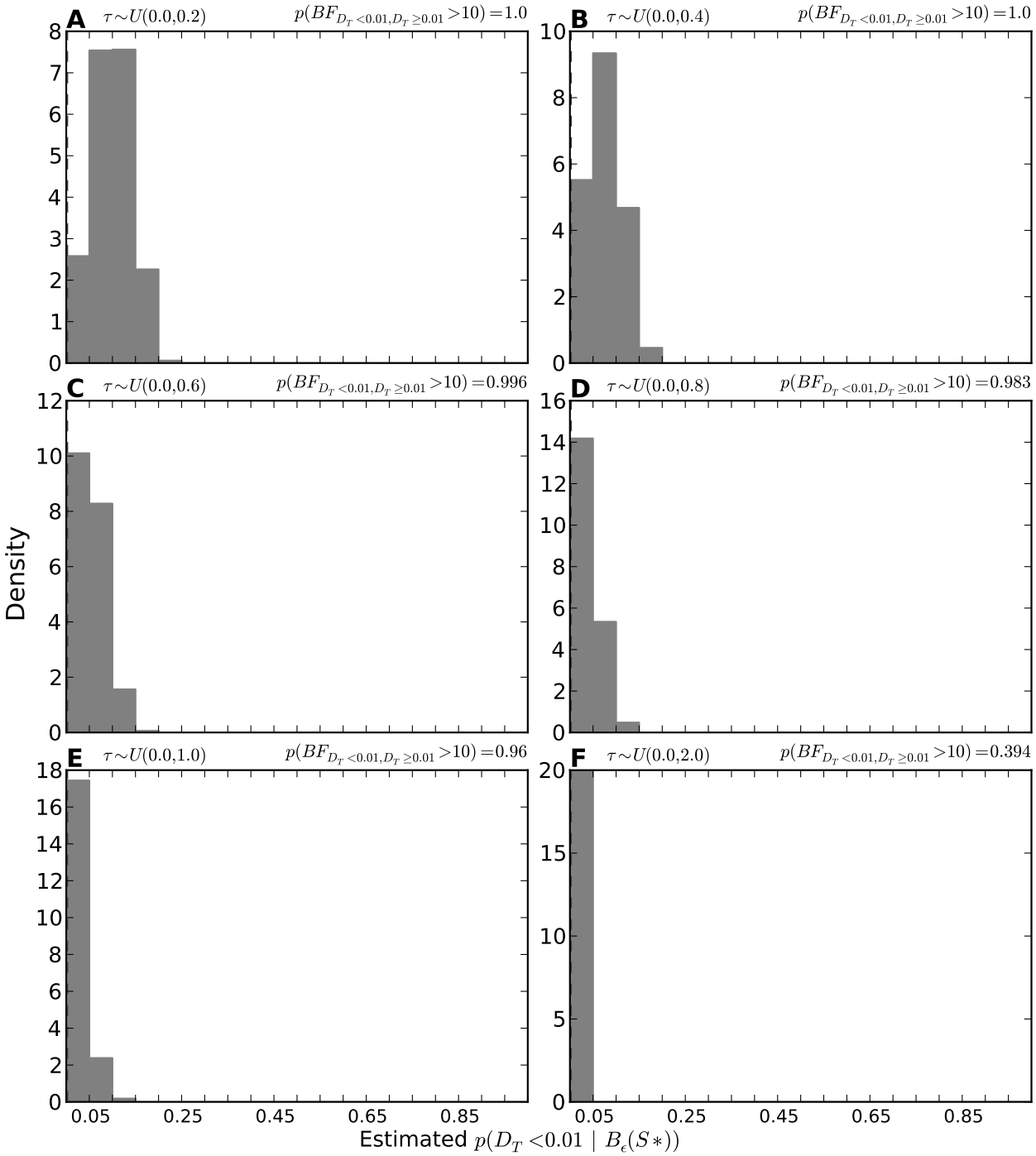

Figure 3.92. The tendency of model $M_{Uniform}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{D_T<0.01,D_T\geq0.01} > 10)$, given for each data model.

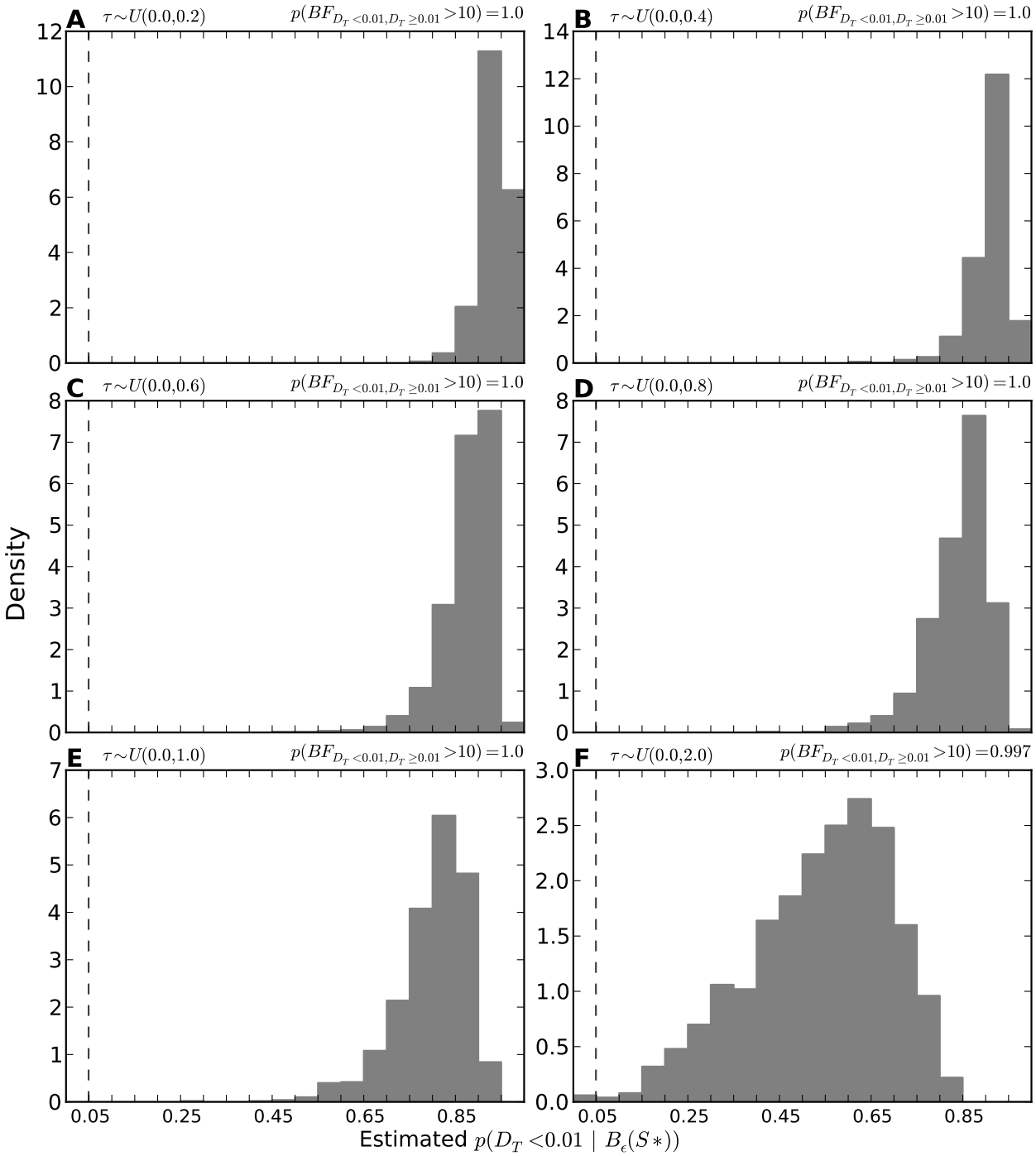

Figure 3.93. The tendency of model $M_{msBayes}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{D_T<0.01, D_T \geq 0.01} > 10)$, given for each data model.

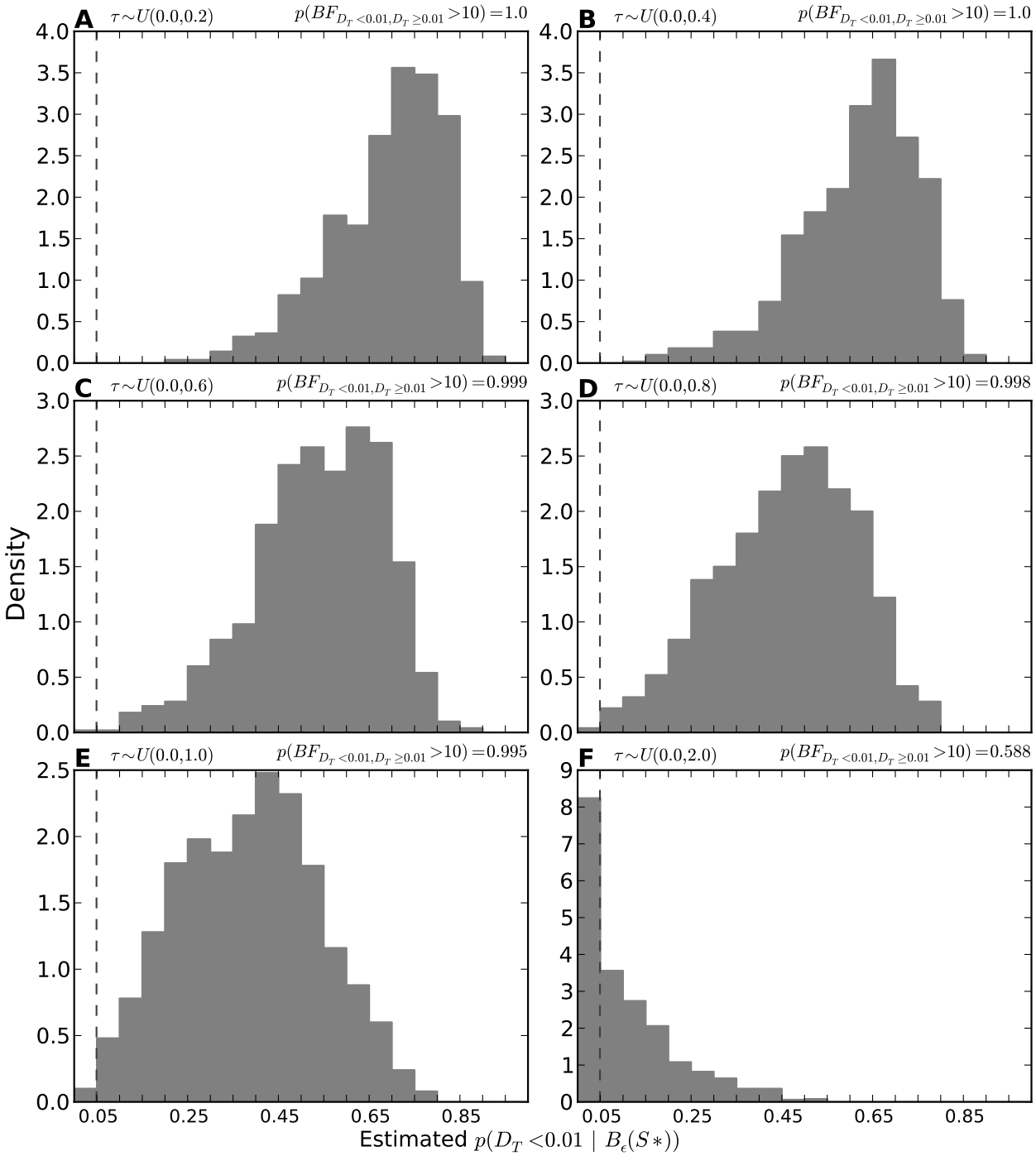

Figure 3.94. The tendency of model $M_{Ushaped}$ to support one divergence event when there is random variation in divergence times as simulated under the series of models $\mathcal{M}_{msBayes}$. The plots illustrate histograms of the estimated posterior probability of the one divergence model, $p(D_T < 0.01|B_\epsilon(\mathbf{S}^*))$, from analyses of 1000 datasets simulated under each of the $\mathcal{M}_{msBayes}$ models, with the the estimated probability of the model strongly supporting one divergence event, $p(BF_{D_T<0.01,D_T\geq0.01} > 10)$, given for each data model.

Chapter 4

Patterns of Colonization and Diversification in the Philippine Archipelago

4.1 Introduction

The Islands of the Philippines may harbor the highest concentration of biodiversity on Earth, and a large proportion of this diversity is comprised of endemic species (Brown, 2009; Heaney, 2007). Over the time this biodiversity was generated, the islands have experienced a long, dynamic, and relatively well-understood geological history (Hall, 1998; Voris, 2000; Yumul et al., 2008). The combination of their biodiversity and dynamic history makes the islands an important model system for investigating how large-scale geological and climatic processes generate and maintain biodiversity.

The Philippines are comprised of both continental and oceanic islands. The former comprise the Palawan Microcontinental Block, which broke away from mainland Asia and were potentially submerged (but see Blackburn et al., 2010; Siler et al., 2012) as they moved southward before re-emerging as islands. The latter comprise the Philippine Mobile Belt

and originated via volcanism much further east in the Pacific as the Philippine Sea Plate moved westward, eventually colliding with the Palawan Microcontinental Block, bringing the islands of the Philippines to their current position just east of the Sunda Shelf (Hall, 1996, 1998; Yumul et al., 2008, 2009).

Our goal is to establish broad-scale temporal patterns of colonization and diversification across a diverse set of vertebrates within the dynamic geological framework that gave rise to the Philippine Islands. We assimilate a large comparative dataset of colonization-time estimates for a diverse set of monophyletic groups of vertebrates that all reside in the Philippines. We use these data to estimate how much of the Philippines' species diversity is explained by the amount of time lineages have had to diversify within the archipelago (i.e., a time-for-diversification effect), and test whether this accounts for a significant proportion of variation in diversity (Wiens et al., 2009; Stephens and Wiens, 2003). Furthermore, we compare patterns of colonization time and diversity across major vertebrate groups and discuss how they might help explain some of the general patterns of biodiversity in the islands.

4.2 Methods

We gathered divergence time estimates for clades of Philippine vertebrates from the literature. We also gathered available sequence datasets that include Philippine taxa for estimating divergence times for additional Philippine groups. We considered all groups of two or more species that are estimated to form a monophyletic group of Philippine species. In particular, we were interested in estimating the time these clades colonized the archipelago. Whenever possible, we approximated the time of colonization of the Philippines using the midpoint between the estimated ages of the node of the most-recent common ancestor (MRCA) of the Philippine clade (i.e., the crown-based node) and the node of the MRCA of the Philippine clade and its most closely related non-Philippine relative(s) (i.e., the stem-based node). We excluded taxa with only a single species in the Philippines due to the inability to establish

a lower bound for the time of invasion for these taxa, other than the present (Wiens et al., 2009). Whenever possible, we also approximated a confidence interval (CI) for each colonization time, which extends from the lower limit of the estimated CI for the age of the crown-based node up to the upper limit of the CI for the age of the stem-based node. When CIs were not available, we used the estimated crown and stem node ages for the bounds. Any exceptions to how we established either the midpoint or CIs are noted below. All of the estimates of invasion times, along with the estimates of species diversity of each clade, are summarized in Table 4.1. We used BEAST (v1.7.5; Drummond and Rambaut, 2007) for all dating analyses, the full details of which, including the XML-formatted analysis files, can be found at https://joaks1@bitbucket.org/joaks1/tfd-philippines.git.

4.2.1 Colonization ages of avian clades

We use node-age estimates of two Philippine clades of *Copsychus* from Lim et al. (2010). This includes the clade comprised of *C. niger* and *C. cebuensis*, and the clade comprised of *C. luzoniensis* and *C. superciliaris* (Table 4.1). We use the age estimates from Lim et al. (2010) that are based on the rate distribution of Nabholz et al. (2009) for sites at the third-codon position within the cytochrome-b gene.

Furthermore, we use the ages of: (1) a three-species clade of the *Irena cyanogastra* group as estimated by Moltesen et al. (2012); (2) a clade comprised of *Hypothymis helenae* and *H. coelestis* as estimated by Fabre et al. (2012) (we use the age estimate and CI for the stem-based node of this clade, because an age estimate for the crown-based node is not provided); (3) a clade comprised of *Terpsiphone unirufa* and *T. cinnamomea* also estimated by Fabre et al. (2012) (the age estimate and CI for both the crown and stem nodes were provided for this group). (4) a nine-species clade of *Aethopyga* as estimated by Hosner et al. (2013) (we use the results of their dating analysis that was calibrated with a 2.4% divergence rate per million years); (5) a clade comprised of *Oriolus steerei* and *O. isabellae* as estimated by Jønsson et al. (2010); (6) a three-species clade of *Rhabdornis* as estimated by Zuccon

et al. (2006) (we use the CI reported for the age of the stem-based node and its midpoint as the colonization-time estimate, because age information for the crown-based node in not provided).

We re-analyzed the ND2 data of Oliveros and Moyle (2010), Oliveros et al. (2012), and Sánchez-González and Moyle (2011) using a strict-clock model with a mutation rate of 1.2% per million years for all three datasets. From our results, we use the estimated ages of (1) a seven-species clade of *Hypsipetes*, (2) a three-species clade of *Robsonius*, and (3) a six-species clade of *Rhipidura* (Table 4.1).

4.2.2 Colonization ages of mammalian clades

We use the age of an eight-species clade of *Crocidura* as estimated by Esselstyn and Brown (2009) using their combined calibration strategy. We assume that the unsampled species *C. grandis* is a member of this clade. We use the age estimates of Jansa et al. (2006) for five clades (labeled A–E by Jansa et al. (2006)) of Murinae. We use the results of their Bayesian analyses of the IRBP locus; we use only point estimates of the stem-based nodes as proxies for invasion-age estimates, as this was the only information provided for these five clades.

4.2.3 Colonization ages of anuran clades

We use the age of a nine-species clade of *Sanguirana* as estimated by Wiens et al. (2009); we use the crown and stem-based nodes as the lower and upper limits of our CI, respectively, because no CIs were provided.

Furthermore, we analyzed the data of Blackburn et al. (2013), Brown and Siler (2013), and Evans et al. (2003) to estimate the ages of a (1) 14-species clade of *Kaloula*, (2) six-species clade of *Hylarana*, and (3) an eight-species clade of *Limnonectes*, respectively (Table 4.1).

4.2.4 Colonization ages of squamate clades

We use the ages of two clades of *Sphenomorphus*-group skinks as estimated by Brown et al. (In press), and a 13-species clade of *Gekko* as estimated by Siler et al. (2012). Furthermore, we re-analyze the following datasets to estimate clade colonization ages: (1) Barley et al. (2013) to obtain an estimate for a 12-species clade of *Eutropis*, (2) Siler et al. (2013) to obtain an estimate for the 9-species clade of *Lycodon*, (3) Siler et al. (2011a) to obtain an estimate of a 41-species clade of *Brachymeles*, (4) Siler et al. (2010) to obtain an estimates of a 12-species clade of *Cyrtodactylus*, (5) Brown et al. (2012b) and Brown et al. (2012a) to estimate the age of a six-species clade of *Pseudogekko*, (6) and unpublished data to estimate the age of an 11-species clade of *Gonocephalus* (Table 4.1). The assumed rates of substitution used in these analyes are summarized in Table 4.2, and full details of the BEAST analyses can be found at https://joaks1@bitbucket.org/joaks1/tfd-philippines.git.

4.2.5 Evaluating the time-for-diversification hypothesis

We plot the clades' estimated times of colonization of the Philippines against diversity, using both the raw and log-transformed number of species. Given that diversity will increase exponentially under a model with a constant positive diversification rate across lineages, log-transforming diversity should linearize the relationship between diversity and time. We use least-squares linear regression to estimate the proportion of the variation in species diversity explained by colonization time, and to test the significance of the relationship.

4.3 Results

Using the raw data, we are unable to reject a null hypothesis of no relationship between time and species diversity with a Type I error rate of 0.05 ($P = 0.068$; Figure 4.1). After log-transforming the number of species for each clade, the relationship is significant ($P = 0.007$; Figure 4.1). However, even when diversity is log-transformed, the time for clade

diversification only explains 23% of the variance in diversity among the groups we analyzed. One interesting pattern that emerges from our results is that all of the squamate groups we analyzed are estimated to have colonized the Philippine Islands before any of the avian, mammalian, or frog groups (Figure 4.1).

4.4 Discussion

Our results suggest a relatively weak relationship between the time of colonization of the islands and present-day species diversity (Figure 4.1). We expected to find a large amount of unexplained variation of diversity. A portion of this unexplained variance is likely due to the stochastic nature of diversification processes. For example, even simple, constant-rate birth-death models, where diversity is a direct function of time, predict a large amount of variance in diversity. Furthermore, there are many deterministic factors (e.g., vagility, range size, and other ecological factors) that also likely account for much of the unexplained variation in diversity.

Perhaps more interesting than the relatively weak time-for-diversification effect, is our finding that of all the vertebrate lineages we examined, all of the squamate taxa are estimated to have arrived in the Philippines before any of the other groups, including birds, mammals, and frogs (Figure 4.1). The colonization times for some of the lizard groups (*Eutropis*, *Gekko*, and *Pseudogekko*) are likely over estimates (Table 4.1) due to long branches connecting the Philippine species to their nearest non-Philippine relatives. Thus, the midpoint between the ages of the crown and stem nodes for these groups is likely a poor estimate of their colonization time. Nonetheless, even the lower limits of the CIs on the time of colonization for these groups (which are based on the age of the crown node) are still older than any of the non-squamate taxa. (Figure 4.1).

This pattern of early squamate invasion is interesting given the geological history of the islands. The islands of the Philippine Mobile Belt are of volcanic origin and formed on the

Philippine Sea Plate further out in the Pacific and subsequently moved westward toward their current location just east of the Sunda Shelf (Hall, 1996, 1998; Yumul et al., 2008, 2009). Thus, many of the islands spent much of their history as remote islands, much farther from any continental sources of biodiversity. Given that lizards are presently a main component of the vertebrate fauna of remote Pacific Islands, it is not surprising to see that they were early colonizers of the Philippines when the islands were much more isolated.

Given the history of the islands and the relative dispersal abilities of the groups, we might expect lizards to be early colonizers relative to mammals and frogs. However, it is somewhat surprising that none of the bird groups we included are early colonizers of the islands given their high vagility and the fact that birds are also common across remote islands of the Pacific. Perhaps this pattern is a reflection of their high dispersal ability driving a higher turnover of avian biodiversity in the islands. This could lead to many of the early avian colonizers being replaced by more recent colonizers as the islands approached the Sunda Shelf. However, this is speculative, and the groups we include in this study are certainly not a random sample of vertebrate biodiversity of the Philippines. Furthermore, our methods for approximating the times of colonization make many simplifying assumptions that are likely violated across these groups. Nonetheless, our goal is to establish a broad picture of the temporal patterns of colonization and diversification using as much information that is currently available. We suspect that some of the general patterns in our results, such as the time-for-diversification effect and early squamate colonization, are relatively robust to biased taxon sampling and model violations.

4.5 Tables

Table 4.1. Estimated times of colonization and species diversity for Philippine vertebrate groups. Colonization-time estimates are based on the midpoint between the crown- and stem-based node of each clade (see text for exceptions). The citation for the paper from which the estimate is based is provided, and we use an asterisk to indicate the taxa for which we re-analyzed data to obtain the time estimates.

Group	Clade	Age (mya)	Lower CI	Upper CI	Species #	Citation
Birds	*Copsychus luzoniensis-superciliaris*	2.575	0.16	9.23	2	Lim et al. (2010)
	Copsychus niger-cebuensis	3.09	0.37	7.91	2	Lim et al. (2010)
	Irena cyanogastra group	2.279	0.941	4.206	3	Moltesen et al. (2012)
	Hypothymis helenae-coelestis	5.9	4.7	7.2	2	Fabre et al. (2012)
	Terpsiphone unirufa-cinnamomea	2.95	1.2	5	2	Fabre et al. (2012)
	Aethopyga	6.34	5.017	7.563	9	Hosner et al. (2013)
	Oriolus steerei-isabellae	4.34	2.641	6.604	2	Jønsson et al. (2010)
	Rhabdornis	19.1	17.8	20.4	3	Zuccon et al. (2006)
	*Hypsipetes**	19.279	12.271	27.513	7	Oliveros and Moyle (2010)
	*Rhipidura**	4.896	2.663	7.425	6	Sánchez-González and Moyle (2011)
	*Robsonius**	13.327	2.054	28.167	3	Oliveros et al. (2012)
Mammals	*Crocidura*	2.5	2.1	3.3	8	Esselstyn and Brown (2009)
	Murinae clade A	3.1	-	-	5	Jansa et al. (2006)
	Murinae clade B	3.1	-	-	3	Jansa et al. (2006)
	Murinae clade C	7.1	-	-	3	Jansa et al. (2006)
	Murinae clade D	15.8	-	-	32	Jansa et al. (2006)
	Murinae clade E	18.7	-	-	14	Jansa et al. (2006)
Squamates	*Eutropis**	107.897	46.047	189.783	12	Barley et al. (2013)
	*Lycodon**	49.366	28.218	72.046	9	Siler et al. (2013)
	*Brachymeles**	58.977	34.363	90.618	41	Siler et al. (2011a)
	*Cyrtodactylus**	43.465	38.220	49.200	12	Siler et al. (2010)
	*Gonocephalus**	54.966	38.845	74.333	11	Unpublished
	*Pseudogekko**	81.780	55.458	115.953	6	Brown et al. (2012a,b)
	Gekko	106.450	65.335	157.860	13	Siler et al. (2012)
	Sphenomorphus clade A	32.852	23.508	43.780	41	Brown et al. (In press)
	Sphenomorphus clade B	24.394	12.592	39.343	4	Brown et al. (In press)
Anurans	*Sanguirana*	21.22	14.4	28.04	9	Wiens et al. (2009)
	*Kaloula**	14.061	11.785	16.429	14	Blackburn et al. (2013)
	*Hylarana**	13.001	10.592	15.643	6	Brown and Siler (2013)
	*Limnonectes**	9.534	8.242	10.777	8	Evans et al. (2003)

Table 4.2. Rates of nucleotide substitution used to calibrate dating analyses. Units are in substitutions per site per million years.

Taxon	Substitution rate
Hypsipetes	0.012
Rhipidura	0.012
Robsonius	0.012
Lycodon	0.01
Eutropis	0.0122
Brachymeles	0.0122
Cyrtodactylus	0.0058
Gonocephalus	0.009
Kaloula	0.0075
Hylarana	0.0075
Pseudogekko	0.0058
Limnonectes	0.0075

4.6 Figures

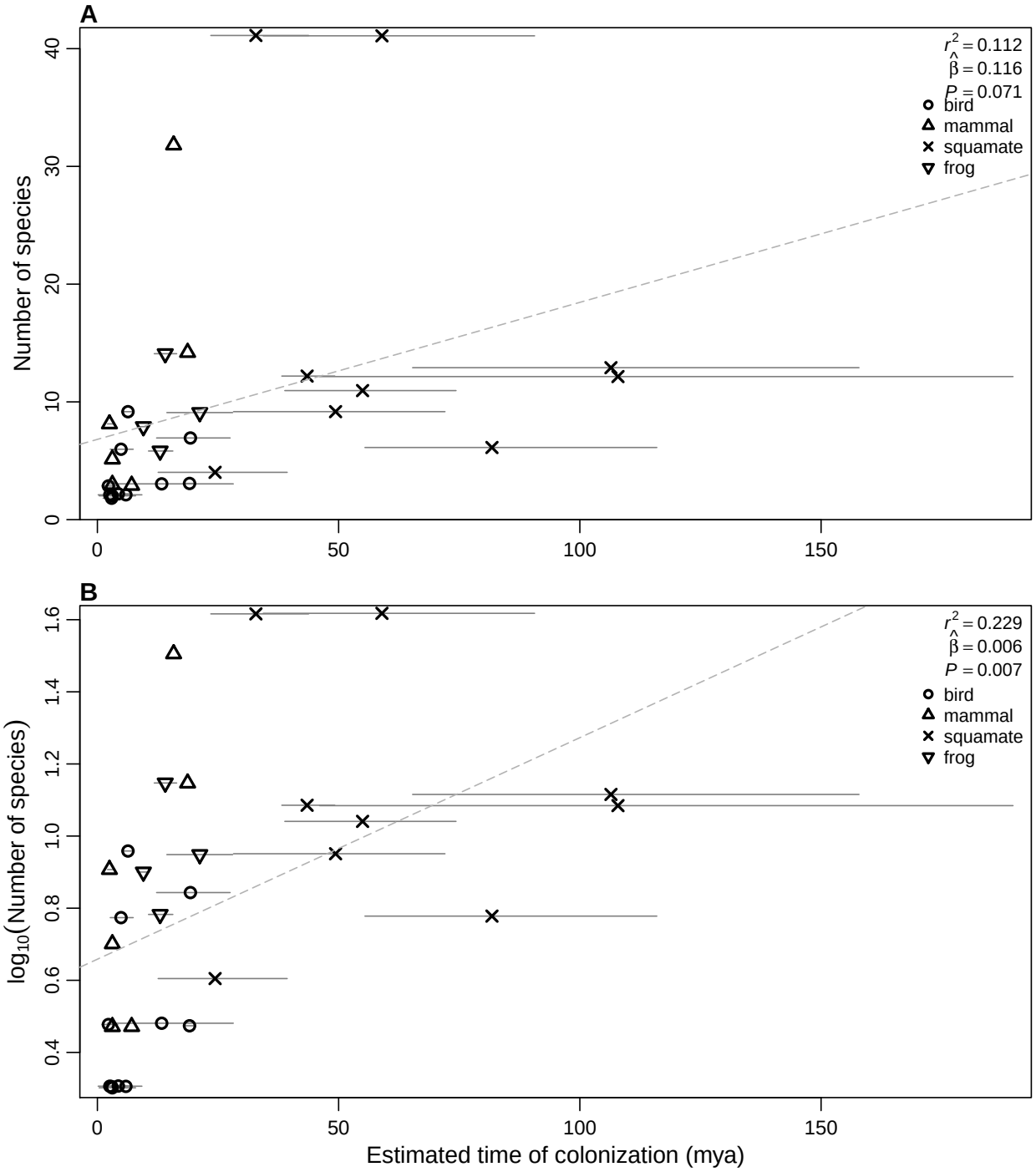

Figure 4.1. The estimated time of lineage colonization plotted against (A) raw and (B) log-transformed present-day number of species. Time of colonization is approximated by the midpoint between the estimated ages of the crown- and stem-based node of each clade, and error bars extend from the lower limit of the confidence interval for the estimated age of the crown-based node up to the upper limit of the confidence interval for the estimated age of the stem-based node (see text for exceptions).

Part V

References

Bibliography

Antoniak, C. E., 1974. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. Annals of Statistics 2:1152–1174.

Ayres, D. L., A. Darling, D. J. Zwickl, P. Beerli, M. T. Holder, P. O. Lewis, J. P. Huelsenbeck, F. Ronquist, D. L. Swofford, M. P. Cummings, A. Rambaut, and M. A. Suchard, 2012. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Systematic Biology 61:170–173.

Barber, B. R. and J. Klicka, 2010. Two pulses of diversification across the Isthmus of Tehuantepec in a montane Mexican bird fauna. Proceedings Of The Royal Society B-Biological Sciences 277:2675–2681.

Barley, A. J., J. White, A. C. Diesmos, and R. M. Brown, 2013. The challenge of species delimitation at the extremes: Diversification without morphological change in Philippine sun skinks. Evolution Pp. 1–17.

Beaumont, M., W. Zhang, and D. J. Balding, 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

Beaumont, M. A., 2010. Approximate Bayesian Computation in Evolution and Ecology. Annual Review of Ecology, Evolution, and Systematics 41:379–406.

Bell, E. T., 1934. Exponential numbers. American Mathematical Monthly 41:411–419.

Bell, R. C., J. B. MacKenzie, M. J. Hickerson, K. L. Chavarria, M. Cunningham, S. Williams, and C. Moritz, 2012. Comparative multi-locus phylogeography confirms multiple vicariance events in co-distributed rainforest frogs. Proceedings Of The Royal Society B-Biological Sciences 279:991–999.

Bertorelle, G., A. Benazzo, and S. Mona, 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. Molecular Ecology 19:2609–2625.

Blackburn, D. C., D. P. Bickford, A. C. Diesmos, D. T. Iskandar, and R. M. Brown, 2010. An ancient origin for the enigmatic flat-headed frogs (Bombinatoridae: *Barbourula*) from the Islands of Southeast Asia. PLoS ONE 5:10.

Blackburn, D. C., C. D. Siler, A. C. Diesmos, J. A. McGuire, D. C. Cannatella, and R. M. Brown, 2013. An adaptive radiation of frogs in a Southeast Asian island archipelago. Evolution 67:2631–2646.

Blum, M. G. B. and O. François, 2009. Non-linear regression models for Approximate Bayesian Computation. Statistics and Computing 20:63–73.

Brown, R. M., 2009. Frogs. Pp. 347–351, *in* R. Gillespie and D. Clague, eds. Encyclopdia of Islands. University of California Press, Berkeley.

Brown, R. M. and A. C. Diesmos, 2002. Application of lineage-based species concept to oceanic island frog populations: The effects of differing taxonomic philosophies on the estimation of Philippine biodiversity. Silliman Journal 42:133–162.

———, 2009. Philippines, biology. Pp. 723–732, *in* R. Gillespie and D. Clague, eds. Encyclopdia of Islands. University of California Press, Berkeley.

Brown, R. M., A. C. Diesmos, and A. C. Alcala, 2008. Philippine amphibian biodiversity is increasing in leaps and bounds. Pp. 82–83, *in* S. N. Stuart, M. Hoffmann, J. S. Chanson, N. A. Cox, R. Berridge, P. Ramani, and B. E. Young, eds. Threatened Amphibians of the

World. Lynx Ediciones, Barcelona, Spain; IUCN—The World Conservation Union, Gland, Switzerland; and Conservation International, Arlington, Virginia, USA.

Brown, R. M. and C. D. Siler, 2013. Spotted stream frog diversification at the Australasian faunal zone interface, mainland versus island comparisons, and a test of the Philippine 'dual-umbilicus' hypothesis. Journal of Biogeography Pp. 1–14.

Brown, R. M., C. D. Siler, I. Das, and Y. Min, 2012a. Testing the phylogenetic affinities of Southeast Asia's rarest geckos: Flap-legged geckos (*Luperosaurus*), flying geckos (*Ptychozoon*) and their relationship to the pan-Asian genus *Gekko*. Molecular Phylogenetics and Evolution 63:915–921.

Brown, R. M., C. D. Siler, L. L. Grismer, I. Das, and J. A. McGuire, 2012b. Phylogeny and cryptic diversification in Southeast Asian flying geckos. Molecular phylogenetics and evolution 65:351–361.

Brown, R. M., C. D. Siler, C. H. Oliveros, J. A. Esselstyn, A. C. Diesmos, P. A. Hosner, C. W. Linkem, A. J. Barley, J. R. Oaks, M. B. Sanguila, L. J. Welton, R. G. Moyle, A. T. Peterson, and A. C. Alcala, In press. Evolutionary processes of diversification in a model island archipelago. Annual Review of Ecology, Evolution, and Systematics .

Brown, R. M. and B. L. Stuart, 2012. Patterns of biodiversity discovery through time: an historical analysis of amphibian species discoveries in the Southeast Asian mainland and island archipelagos. Pp. 348–389, *in* D. J. Gower, K. G. Johnson, J. E. Richardson, B. R. Rosen, L. Rüber, and S. T. Williams, eds. Biotic Evolution and Environmental Change in Southeast Asia. Cambridge University Press.

Burbrink, F. T., R. Lawson, and J. B. Slowinski, 2000. Mitochondrial DNA phylogeography of the polytypic North American Rat Snake (*Elaphe obsoleta*): A critique of the subspecies concept. Evolution 54:2107–2118.

Carlin, B. P. and A. E. Gelfand, 1990. Approaches for empirical Bayes confidence intervals. Journal of the American Statistical Association 85:105–114.

Carnaval, A. C., M. J. Hickerson, C. F. B. Haddad, M. T. Rodrigues, and C. Moritz, 2009. Stability Predicts Genetic Diversity in the Brazilian Atlantic Forest Hotspot. Science 323:785–789.

Chan, L. M., J. L. Brown, and A. D. Yoder, 2011. Integrating statistical genetic and geospatial methods brings new power to phylogeography. Molecular Phylogenetics and Evolution 59:523–537.

Csilléry, K., M. G. B. Blum, O. E. Gaggiotti, and O. François, 2010. Approximate Bayesian Computation (ABC) in practice. Trends In Ecology and Evolution 25:410–418.

Daza, J. M., T. A. Castoe, and C. L. Parkinson, 2010. Using regional comparative phylogeographic data from snake lineages to infer historical processes in Middle America. Ecography 33:343–354.

Dickerson, R. E., 1928. Distribution of life in the Philippines. Philippine Bureau of Science, Manila, Philippines.

Drummond, A. J. and A. Rambaut, 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Edgar, R. C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792–1797.

Efron, B., 2013. Empirical bayes modeling, computation, and accuracy. Manuscript AMS 2010 subject classifications: Primary 62C10; secondary 62-07, 62P10.

Esselstyn, J. A. and R. M. Brown, 2009. The role of repeated sea-level fluctuations in the generation of shrew (Soricidae: *Crocidura*) diversity in the Philippine Archipelago. Molecular Phylogenetics and Evolution 53:171–181.

Esselstyn, J. A., B. J. Evans, J. L. Sedlock, F. A. A. Khan, and L. R. Heaney, 2012. Single-locus species delimitation: A test of the mixed Yule-coalescent model, with an empirical application to Philippine round-leaf bats. Proceedings Of The Royal Society B-Biological Sciences 279:3678–3686.

Esselstyn, J. A., R. M. Timm, and R. M. Brown, 2009. Do geological or climatic processes drive speciation in dynamic archipelagos? the tempo and mode of diversification in Southeast Asian shrews. Evolution 63:2595–2610.

Evans, B., R. Brown, J. Mcguire, J. Supriatna, N. Andayani, A. Diesmos, D. Iskandar, D. Melnick, and D. Cannatella, 2003. Phylogenetics of fanged frogs: Testing biogeographical hypotheses at the interface of the Asian and Australian faunal zones. Systematic Biology 52:794–819.

Fabre, P.-H., M. Irestedt, J. Fjeldså, R. Bristol, J. J. Groombridge, M. Irham, and K. A. Jønsson, 2012. Dynamic colonization exchanges between continents and islands drive diversification in paradise-flycatchers (*Terpsiphone*, Monarchidae). Journal of Biogeography 39:1900–1918.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. Journal of Molecular Evolution 17:368–376.

Ferguson, T. S., 1973. A Bayesian Analysis of Some Nonparametric Problems. The Annals of Statistics 1:209–230.

Green, P. J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711–732.

Hall, R., 1996. Reconstructing Cenozoic SE Asia. Pp. 153–184, *in* R. Hall and D. Blundell, eds. Tectonic Evolution of Southeast Asia. Geological Society of London Special Publication, London.

———, 1998. The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. Pp. 99–131, *in* R. Hall and J. D. Holloway, eds. Biogeography and geological evolution of SE Asia. Backhuys Publishers, Leiden.

Haq, B. U., J. Hardenbol, and P. R. Vail, 1987. Chronology of fluctuating sea levels since the Triassic. Science 235:1156–1167.

Hasegawa, M., H. Kishino, and T.-a. Yano, 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22:160–174.

Heaney, L. R., 1985. Zoogeographic evidence for middle and late pleistocene land bridges to the philippine islands. Mod Quatern Res SE Asia 9:127–144.

———, 1986. Biogeography of mammals in se asia: estimates of rates of colonization, extinction and speciation. Biological Journal of the Linnean Society 28:127–165.

———, 2007. Is a new paradigm emerging for oceanic island biogeography? Journal of Biogeography 34:753–757.

Heaney, L. R. and J. C. Regalado, Jr., 1998. Vanishing treasures of the Philippine rain forest. Field Museum, Chicago, Illinois.

Hickerson, M. J., E. A. Stahl, and H. A. Lessios, 2006. Test for simultaneous divergence using approximate Bayesian computation. Evolution 60:2435–2453.

Hickerson, M. J., G. N. Stone, K. Lohse, T. C. Demos, X. Xie, C. Landerer, and N. Takebayashi, 2013. Recommendations for using msbayes to incorporate uncertainty in selecting an ABC model prior: A response to Oaks et al. Evolution Pp. n/a–n/a.

Hosner, P. A., A. S. Nyári, and R. G. Moyle, 2013. Water barriers and intra-island isolation contribute to diversification in the insular *Aethopyga* sunbirds (Aves: Nectariniidae). Journal of Biogeography 40:1094–1106.

Huang, W., N. Takebayashi, Y. Qi, and M. J. Hickerson, 2011. MTML-msBayes: Approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. BMC Bioinformatics 12:1.

Hudson, R. R., 1990. Gene genealogies and the coalescent process. Pp. 1–44, *in* D. Futuyma and J. Antonovics, eds. Oxford Surveys in Evolutionary Biology, vol. 7. Oxford University Press.

Huelsenbeck, J. P. and B. Rannala, 2004a. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Systematic Biology 53:904–913.

———, 2004b. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Systematic Biology 53:904–913.

Hwang, J. T. G., J. Qiu, and Z. Zhao, 2009. Empirical Bayes confidence intervals shrinking both means and variances. Journal of the Royal Statistical Society Series B-Statistical Methodology 71:265–285.

Inger, R. F., 1954. Systematics and zoogeography of Philippine Amphibia. Fieldiana 33:182–531.

Jansa, S. A., F. K. Barker, and L. R. Heaney, 2006. The pattern and timing of diversification of Philippine endemic rodents: Evidence from mitochondrial and nuclear gene sequences. Systematic Biology 55:73–88.

Jeffreys, H., 1935. Some tests of significance, treated by the theory of probability. Proceedings of the Cambridge Philosophical Society 31:203–222.

———, 1961. Theory of Probability. 3rd ed. Oxford University Press, Oxford, U.K.

Jønsson, K. A., R. C. K. Bowie, R. G. Moyle, M. Irestedt, L. Christidis, J. A. Norman,

and J. Fjeldså, 2010. Phylogeny and biogeography of Oriolidae (Aves: Passeriformes). Ecography 33:232–241.

Joyce, P. and P. Marjoram, 2008. Approximately sufficient statistics and Bayesian computation. Statistical Applications in Genetics and Molecular Biology 7:1–16.

Kingman, J. F. C., 1982. The coalescent. Stochastic processes and their applications 13:235–248.

Laird, N. M. and T. A. Louis, 1987. Empirical Bayes confidence intervals based on bootstrap samples. Journal of the American Statistical Association 82:739–750.

———, 1989. Empirical Bayes confidence intervals for a series of related experiments. Biometrics 45:481–495.

Lawson, L. P., 2010. The discordance of diversification: evolution in the tropical-montane frogs of the Eastern Arc Mountains of Tanzania. Molecular Ecology 19:4046–4060.

Leaché, A. D., S. C. Crews, and M. J. Hickerson, 2007. Two waves of diversification in mammals and reptiles of Baja California revealed by hierarchical Bayesian analysis. Biology Letters 3:646–650.

Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard, 2009. Bayesian phylogeography finds its roots. PLoS Computational Biology 5:e1000520.

Leuenberger, C. and D. Wegmann, 2010. Bayesian computation and model selection without likelihoods. Genetics 184:243–252.

Lim, H. C., F. Zou, S. S. Taylor, B. D. Marks, R. G. Moyle, G. Voelker, and F. H. Sheldon, 2010. Phylogeny of magpie-robins and shamas (Aves: Turdidae: *Copsychus* and *Trichixos*): implications for island biogeography in Southeast Asia. Journal of Biogeography 37:1894–1906.

Lindley, D. V., 1957. A statistical paradox. Biometrika 44:187–192.

Linkem, C. W. and R. M. Brown, 2013. Systematic revision of the *Parvoscincus decipiens* (Boulenger, 1894) complex of Philippine forest skinks (Squamata: Scincidae: Lygosominae) with descriptions of seven new species. Zootaxa 3700:501–533.

Linkem, C. W., A. C. Diesmos, and R. M. Brown, 2011. Molecular systematics of the Philippine forest skinks (Squamata: Scincidae: *Sphenomorphus*): testing morphological hypotheses of interspecific relationships. Zoological Journal of the Linnean Society 163:1217–1243.

Linkem, C. W., K. M. Hesed, A. C. Diesmos, and R. M. Brown, 2010. Species boundaries and cryptic lineage diversity in a Philippine forest skink complex (Reptilia; Squamata; Scincidae: Lygosominae). Molecular Phylogenetics and Evolution 56:572–585.

Malenfant, J., 2011. Finite, closed-form expressions for the partition function and for Euler, Bernoulli, and Stirling numbers. arXiv.org arXiv:1103.1585v6 [math.NT].

Miller, K. G., M. A. Kominz, J. V. Browning, J. D. Wright, G. S. Mountain, M. E. Katz, P. J. Sugarman, B. S. Cramer, N. Christie-Blick, and S. F. Pekar, 2005. The Phanerozoic record of global sea-level change. Science 310:1293–1298.

Moltesen, M., M. Irestedt, J. Fjeldså, P. G. P. Ericson, and K. A. Jønsson, 2012. Molecular phylogeny of Chloropseidae and Irenidae—cryptic species and biogeography. Molecular phylogenetics and evolution 65:903–14.

Morris, C. N., 1983. Parametric empirical bayes inference: Theory and applications. Journal of the American Statistical Association 78:47–55.

Nabholz, B., S. Glémin, and N. Galtier, 2009. The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. BMC Evolutionary Biology 9:54.

Nei, M. and W.-H. Li, 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences 76:5269–5273.

Oaks, J. R., J. Sukumaran, J. A. Esselstyn, C. W. Linkem, C. D. Siler, M. T. Holder, and R. M. Brown, 2013. Evidence for climate-driven diversification? a caution for interpreting ABC inferences of simultaneous historical events. Evolution 67:991–1010.

Oliveros, C. H. and R. G. Moyle, 2010. Origin and diversification of Philippine bulbuls. Molecular Phylogenetics and Evolution 54:822–832.

Oliveros, C. H., S. Reddy, and R. G. Moyle, 2012. The phylogenetic position of some Philippine "babblers" spans the muscicapoid and sylvioid bird radiations. Molecular phylogenetics and evolution 65:799–804.

Plouviez, S., T. M. Shank, B. Faure, C. Daguin-Thiebaut, F. Viard, F. H. Lallier, and D. Jollivet, 2009. Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South. Molecular Ecology 18:3903–3917.

Posada, D. and K. A. Crandall, 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.

Robert, C. P., J.-M. Cornuet, J.-M. Marin, and N. S. Pillai, 2011. Lack of confidence in approximate Bayesian computation model choice. Proceedings of the National Academy of Sciences 108:15112–15117.

Roberts, T. E., 2006a. History, ocean channels, and distance determine phylogeographic patterns in three widespread Philippine fruit bats (Pteropodidae). Molecular Ecology 15:2183–2199.

———, 2006b. Multiple levels of allopatric divergence in the endemic Philippine fruit bat

Haplonycteris fischeri (Pteropodidae). Biological Journal of the Linnean Society 88:329–349.

Rohling, E. J., M. Fenton, F. J. Jorissen, P. Bertrand, G. Ganssen, and J. P. Caulet, 1998. Magnitudes of sea-level lowstands of the past 500,000 years. Nature 394:162–165.

Sánchez-González, L. A. and R. G. Moyle, 2011. Molecular systematics and species limits in the Philippine fantails (Aves: Rhipidura). Molecular Phylogenetics and Evolution 61:290–299.

Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6:461–464.

Siddall, M., E. J. Rohling, A. Almogi-Labin, C. Hemleben, D. Meischner, I. Schmelzer, and D. A. Smeed, 2003. Sea-level fluctuations during the last glacial cycle. Nature 423:853–858.

Siler, C. D. and R. M. Brown, 2010. Phylogeny-based species delimitation in Philippine slender skinks (Reptilia: Squamata: Scincidae: *Brachymeles*): Taxonomic revision of pentadactyl species groups and description of three new species. Herpetological Monographs 24:1–54.

Siler, C. D., A. C. Diesmos, A. C. Alcala, and R. M. Brown, 2011a. Phylogeny of Philippine slender skinks (Scincidae: *Brachymeles*) reveals underestimated species diversity, complex biogeographical relationships, and cryptic patterns of lineage diversification. Molecular Phylogenetics and Evolution 59:53–65.

Siler, C. D., A. M. Fuiten, R. M. Jones, A. C. Alcala, and R. M. Brown, 2011b. Phylogeny-based species delimitation in Philippine slender skinks (Reptilia: Squamata: Scincidae) II: Taxonomic revision of *Brachymeles samarensis* and description of five new species. Herpetological Monographs 25:76–112.

Siler, C. D., J. R. Oaks, J. A. Esselstyn, A. C. Diesmos, and R. M. Brown, 2010. Phylogeny and biogeography of Philippine bent-toed geckos (Gekkonidae: *Cyrtodactylus*) contradict

a prevailing model of Pleistocene diversification. Molecular Phylogenetics and Evolution 55:699–710.

Siler, C. D., J. R. Oaks, L. J. Welton, C. W. Linkem, J. Swab, A. C. Diesmos, and R. M. Brown, 2012. Did geckos ride the Palawan raft to the Philippines? Journal of Biogeography 39:1217–1234.

Siler, C. D., C. H. Oliveros, A. Santanen, and R. M. Brown, 2013. Multilocus phylogeny reveals unexpected diversification patterns in Asian wolf snakes (genus *Lycodon*). Zoologica Scripta 42:262–277.

Sloan, N. J. A., 2011a. The On-Line Encyclopedia of Integer Sequences, Sequence A000041. Published electronically at http://oeis.org.

———, 2011b. The On-Line Encyclopedia of Integer Sequences, Sequence A008284. Published electronically at http://oeis.org.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stamatakis, A., F. Blagojevic, D. Nikolopoulos, and C. Antonopoulos, 2007. Exploring new search algorithms and hardware for phylogenetics: RAxML meets the IBM cell. Journal of VLSI Signal Processing 48:271–286.

Stephens, P. R. and J. J. Wiens, 2003. Explaining species richness from continents to communities: the time-for-speciation effect in emydid turtles. The American naturalist 161:112–128.

Stone, G. N., K. Lohse, J. A. Nicholls, P. Fuentes-Utrilla, F. Sinclair, K. Schönrogge, G. Csóka, G. Melika, J.-L. Nieves-Aldrey, J. Pujade-Villar, M. Tavakoli, R. R. Askew, and M. J. Hickerson, 2012. Reconstructing community assembly in time and space reveals enemy escape in a Western Palearctic insect community. Current Biology 22:532–537.

Sukumaran, J., 2012. Geographies and Genealogies: Phylogeographic Simulation and Bayesian Approaches to Statistical Phylogeographic Model Selection. Ph.D. thesis, University of Kansas, Lawrence, Kansas, USA.

Sukumaran, J. and M. T. Holder, 2010. DendroPy: A Python library for phylogenetic computing. Bioinformatics 26:1569–1571.

Swofford, D. L., 2003. PAUP*, phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, MA.

Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460.

———, 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Takahata, N. and M. Nei, 1985. Gene genealogy and variance of interpopulational nucleotide differences. Genetics 110:325–344.

Topp, C. M. and K. Winker, 2008. Genetic patterns of differentiation among five landbird species from the Queen Charlotte Islands, British Columbia. Auk 125:461–472.

Voje, K. L., C. Hemp, Ø. Flagstad, G.-P. Saetre, and N. C. Stenseth, 2009. Climatic change as an engine for speciation in flightless Orthoptera species inhabiting African mountains. Molecular Ecology 18:93–108.

Voris, H. K., 2000. Maps of the Pleistocene sea levels in South East Asia: Shorelines, river systems, time durations. Journal of Biogeography 27:1153–1167.

Wakeley, J., 2009. Coalescent Theory: An Introduction. Roberts and Company Publishers, Greenwood Village, Colorado, USA.

Watterson, G. A., 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology 7:256–276.

Wegmann, D., C. Leuenberger, and L. Excoffier, 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. Genetics 182:1207–1218.

Wegmann, D., C. Leuenberger, S. Neuenschwander, and L. Excoffier, 2010. ABCtoolbox: a versatile toolkit for approximate bayesian computations. BMC Bioinformatics 11:116.

Welton, L. J., C. D. Siler, C. W. Linkem, A. C. Diesmos, and R. M. Brown, 2010. Philippine bent-toed geckos of the *Cyrtodactylus agusanensis* complex: Multilocus phylogeny, morphological diversity, and descriptions of three new species. Herpetological Monographs 24:55–85.

Wiens, J. J., J. Sukumaran, R. A. Pyron, and R. M. Brown, 2009. Evolutionary and biogeographic origins of high tropical diversity in old world frogs (Ranidae). Evolution 63:1217–1231.

Yumul, G., C. Dimalanta, V. Maglambayan, and E. Marquez, 2008. Tectonic setting of a composite terrane: A review of the Philippine island arc system. Geosciences Journal 12:7–17.

Yumul, G., C. Dimalanta, K. Queano, and E. Marquez, 2009. Philippines, geology. *in* R. Gillespie and D. Clague, eds. Encyclopdia of Islands. University of California Press, Berkeley.

Zuccon, D., A. Cibois, E. Pasquet, and P. G. P. Ericson, 2006. Nuclear and mitochondrial sequence data reveal the major lineages of starlings, mynas and related taxa. Molecular Phylogenetics and Evolution 41:333–344.

Part VI

Appendices

Appendix

5.7 Python functions for calculating integer partition

```
def integer_partition(n):
    """
    Takes an integer (n) and returns the total number of partitions of n, and
    a dictionary of the number of partitions of n for each number of categories
    (k).
    """
    part_nums = {}
    for k in range(1, n + 1):
        part_nums[k] = A(n, k)
    return sum(part_nums.values()), part_nums


def A(n, k):
    """
    Takes two integers, (n) and the number of categories (k) to partition n,
    and returns the number of partitions of n into k categories.
    """
    if k == 1 or n == k:
        return 1
    if k > n or k < 1:
        return 0
    return A(n - 1, k - 1) + A(n - k, k)
```

5.8 Ruling out no divergence

We investigated whether `msBayes` could distinguish our empirical data from cases of no divergence (i.e., panmixia). We simulated 1000 datasets with $\theta_D \sim U(0.0001, 0.1)$, $\theta_A \sim U(0.0001, 0.05)$, and τ fixed to zero. We then analyzed each of the 1000 replicates using 5×10^6 samples from $f(\Theta)$ where $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$. For each analysis we used $\mathbb{S}_{stats}$ and ABC_{GLM}. We compared the observed estimate of $E(\tau)$ to the distribution of simulated estimates, to infer the probability of our empirical data under the hypothesis of panmixia.

The results reject a model of panmixia. Only one of the 1000 simulations under panmixia yielded an estimate of $E(\tau)$ greater than the empirical estimate; most of the simulated estimates were much less than the empirical estimate (Fig. S5.7).

5.9 Correction to `msBayes` Source Code

When using `msBayes`, one must specify a_θ, b_{θ_D}, and b_{θ_A} to control the uniform prior distributions on the θ parameters: $\theta_D \sim U(a_\theta, b_{\theta_D})$ and $\theta_A \sim U(a_\theta, b_{\theta_A})$. b_{θ_A} is specified by a scalar (s) of b_{θ_D}. Thus, we can write the prior on θ_A as

$$\theta_A \sim U(a_\theta, b_{\theta_D} \times s) \tag{5.1}$$

However, on lines 423 and 424 of the `msprior.c` file distributed with version 20100519 of `msBayes`, where θ_A is actually drawn from the prior, the lower bound (a_θ) of the uniform prior on θ_A is mistakenly hard-coded as 0.01:

```
Nanc = gsl_ran_flat (gBaseRand, 0.01,
            gParam.upperAncPopSize * gParam.upperTheta);
```

Thus, the actual prior distribution implemented in `msBayes` is $\theta_A \sim U(0.01, b_{\theta_A})$ *not* $\theta_A \sim U(a_\theta, b_{\theta_A})$. We corrected this part of the code to

```
Nanc = gsl_ran_flat (gBaseRand, gParam.lowerTheta,
            gParam.upperAncPopSize * gParam.upperTheta);
```

for all of the analyses presented in the main text and all of the tables and figures up until Table S5.6 and Figure S5.14.

Before discovering this error, we explored a large number of empirical and simulation-based analyses using alternative summary statistics ($\mathbb{S}_{stats}$ and $\mathbb{S}_{PLS}$), regression methods (ABC_{LLR} and ABC_{GLM}), and prior sample sizes (2×10^6, 5×10^6, and 10^7). All of these analyses are valid as long as the lower bound on the uniform prior on θ_A is recognized as 0.01, rather than the bound specified in our analyses. The results of all the analyses prior to correcting the bug are presented in Figures S5.14–5.42 and Tables S5.6 and S5.7 for two purposes. First, the results of the simulation-based analyses demonstrate that the bias of `msBayes` towards simpler divergence models is consistent across all the combinations of methods we explored. Given the bug has little affect on the simulation-based analyses (because the simulated data and prior samples are both drawn from the incorrect prior on θ_A), these results demonstrate that our simulation results presented in the main text that use $\mathbb{S}_{stats}$, ABC_{GLM}, and 5×10^6 prior samples are representative of a wide range of ABC methods. Second, the results from the empirical data demonstrate sensitivity to the bug correction.

5.10 Results Prior to Bug Correction

5.10.1 Simulation-based assessment of "simultaneous"

Accuracy and precision of estimates

The precision of the ABC estimates of $E(\tau)$ and Ω was relatively low, especially when the simulated divergence times are more recent (i.e., lower τ_{max}), regardless of the post-sampling adjustment method or summary statistics used (Figs. S5.14–5.21). The PLS components ($\mathbb{S}_{PLS}$) and `msBayes` summary statistics ($\mathbb{S}_{stats}$) produced similar results after adjustment by ABC_{LLR} and ABC_{GLM} (Figs. S5.14–5.21). `msBayes` was less accurate and precise in estimating Ω (Figs. S5.18–5.21) than $E(\tau)$ (Figs. S5.14–5.17). For $\hat{E(\tau)}$, $\mathbb{S}_{stats}$ and $\mathbb{S}_{PLS}$ performed similarly, as did ABC_{GLM} and ABC_{LLR} (Fig. S5.14–5.17). All methods struggle with both accuracy and precision of Ω estimates, and performance varies depending on the distribution placed on τ (Figs. S5.18–5.21). ABC_{GLM} tends to have more frequent large overestimations of Ω (Fig. S5.19 and S5.21) than ABC_{LLR} (Fig. S5.18 and S5.20), but both methods have low precision regardless of the distribution placed on τ. Likewise, using $\mathbb{S}_{PLS}$ tends to produce more large overestimations of Ω (Fig. S5.20 and S5.21) than $\mathbb{S}_{stats}$ (Fig. S5.18 and S5.19). All combinations of summary statistics and adjustment methods are inaccurate for estimating Ω across the τ_{max} we simulated.

Power of `msBayes`

The power of `msBayes` to detect random variation in divergence times among the simulated population pairs is low regardless of the post-sampling adjustment method, summary statistics, or prior settings used (Figs. S5.22–5.28, S5.33–5.35, and S5.38–5.41). `msBayes` consistently infers highly clustered divergence times across all the τ_{max} we simulated, and infers the extreme case of $\hat{\Psi} = 1$ in more than 5% of replicates until τ_{max} is well above $4N_C$ generations, regardless of the ABC methods used (Figs. S5.22–5.24, S5.33, S5.38, and S5.39). Note, the multinomial logistic regression adjustment of Ψ provided with `msBayes` failed for approximately 2-30% of the simulations, and the failure rate was positively correlated with the number of posterior samples with $\Psi = 1$. Thus, the simulation results based on the successfully adjusted replicates are biased towards larger Ψ. Because of this, we do not show plots of all these results. However, we do include one figure of the logistic-regression-adjusted Ψ estimates (Fig. S5.23) to demonstrate that despite the bias introduced by the failed replicates, the results are still very similar to the ABC_{GLM}-adjusted results (Fig. S5.24).

Based on estimates of Ω, `msBayes` infers one divergence event (i.e., $\hat{\Omega} < 0.01$) for more than 5% of the replicates when τ_{max} is less than 1.4, 0.9, 1.3, and 0.7 coalescent units when using ABC_{GLM} with $\mathbb{S}_{PLS}$, ABC_{GLM} with $\mathbb{S}_{stats}$, ABC_{LLR} with $\mathbb{S}_{PLS}$, and ABC_{LLR} with $\mathbb{S}_{stats}$, respectively (Figs. S5.28, S5.26, S5.27, and S5.25).

The accuracy, precision, and power of `msBayes` was very similar when we used 10^7 (Figs. S5.29–5.35) and 2×10^6 samples from the prior (Figs. S5.14, S5.15, S5.18, S5.19, S5.22, S5.25 and S5.26).

5.10.2 Empirical results and bug sensitivity

Our results show that `msBayes` analyses of the empirical data are sensitive to the source-code error associated with the prior distribution on θ_A (Tables 1.2, S5.6, and S5.7). Estimates of Ω and the posterior probability of the inferred Ψ change as much as two-fold, and the estimates of $E(\tau)$ change by as much as 46% (Table 5.7).

5.11 Tables

Table S5.3. Summary of the data collected from the 22 population pairs from the Greater Mindanao and Negros-Panay Pleistocene Aggregate Island Complexes (PAICs). On the map, dark gray represents the extent of land at current sea-level, and light gray is the extent when sea-level is 100m lower (data from ETOPO2; `http://www.ngdc.noaa.gov/mgg/global/etopo2.html`).

Species	Gene	bp	n_1	n_2
Mammals				
Crocidura beatus	cytb/ND2	2037	12	11
Crocidura negrina-panayensis	cytb/ND2	2037	12	6
Hipposideros obscurus	ND2	1044	19	9
Hipposideros pygmaeus	ND2	1044	3	12
Cynopterus brachyotis	ND2	600	20	8
Cynopterus brachyotis	ND2	600	8	14
Haplonycteris fischeri	ND2	564	29	8
Haplonycteris fischeri	ND2	564	9	21
Macroglossus minimus	ND2	576	19	4
Macroglossus minimus	ND2	576	8	10
Ptenochirus jagori	ND2	681	4	7
Ptenochirus jagori	ND2	681	8	8
Ptenochirus minor	ND2	681	30	9
Squamates				
Cyrtodactylus gubaot-sumuroi	ND2	1420	29	6
Cyrtodactylus annulatus	ND2	1420	14	3
Cyrtodactylus philippinicus	ND2	1420	6	14
Gekko mindorensis	ND2	1408	8	11
Insulasaurus arborens	ND2	1035	22	10
Pinoyscincus jagori	ND2	1035	8	8
Dendrelaphis marenae	cytb	1067	6	6
Anurans				
Limnonectes leytensis	12S/16S	2433	4	2
Limnonectes magnus	12S/16S	2433	2	3

Map

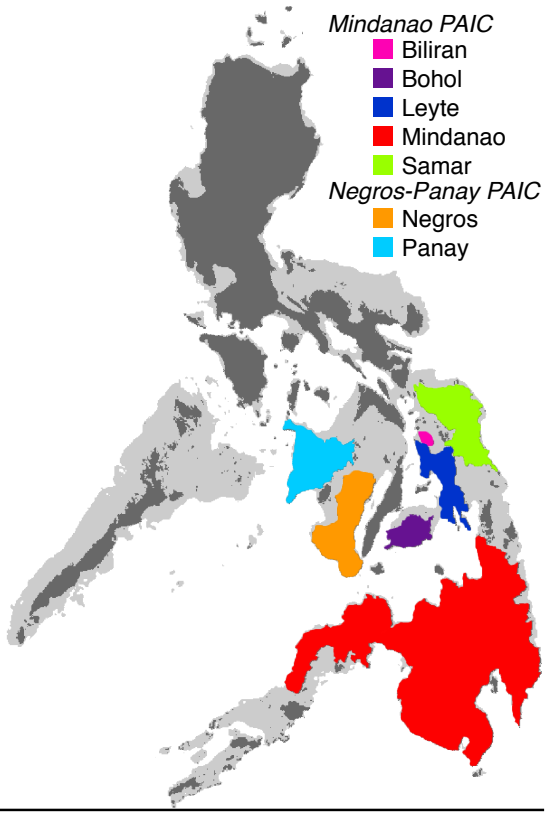

Table S5.4. Estimates of the population mutation rate (θ) from different island populations of the same species used in the `msBayes` analyses and/or closely related species.

Species	**Island**	n	θ_W	π
Crocidura beatus	Camiguin	16	0.0022	0.0021
Crocidura palawanensis	Palawan	23	0.0064	0.0046
Cynopterus brachyotis	Siquijor	7	0.0034	0.0029
Cynopterus brachyotis	Luzon	30	0.0118	0.0077
Cynopterus brachyotis	Sibuyan	8	0.0077	0.0053
Cynopterus brachyotis	Camiguin	8	0.0058	0.0070
Cynopterus brachyotis	Polillo	7	0.0061	0.0060
Cyrtodactylus mamanwa	Dinagat	19	0.0066	0.0041
Cyrtodactylus philippinicus	Luzon	10	0.0037	0.0038
Cyrtodactylus philippinicus	Polillo	14	0.0146	0.0120
Cyrtodactylus philippinicus	Camiguin Norte	10	0.0050	0.0031
Dendrelaphis marenae	Guimaras	6	0.0012	0.0011
Dendrelaphis marenae	Luzon	6	0.0049	0.0042
Gekko gigante	Gigantes	11	0.0019	0.0014
Haplonycteris fischeri	Catanduanes	13	0.0029	0.0024
Haplonycteris fischeri	Mindoro	8	0.0048	0.0034
Haplonycteris fischeri	Sibuyan	20	0.0020	0.0015
Hipposideros obscurus	Luzon	21	0.0100	0.0137
Hipposideros pygmaeus	Polillo	5	0.0049	0.0043
Limnonectes woodworthi	Luzon	9	0.0067	0.0056
Macroglossus minimus	Sibuyan	8	0.0181	0.0162
Macroglossus minimus	Camiguin	8	0.0040	0.0026
Macroglossus minimus	Siquijor	8	0.0074	0.0053
Macroglossus minimus	Luzon	11	0.0166	0.0171
Pinoyscincus abdictus	Polillo	12	0.0042	0.0028
Ptenochirus jagori	Luzon	24	0.0149	0.0110
Ptenochirus jagori	Catanduanes	8	0.0130	0.0102
Ptenochirus jagori	Camiguin	8	0.0074	0.0066
Ptenochirus jagori	Polillo	8	0.0085	0.0089
Ptenochirus jagori	Siquijor	8	0.0068	0.0068

Table S5.5. Estimates of the population mutation rate (θ) for the island populations used in the `msBayes` analyses.

Species	**Island**	n	θ_W	π
Crocidura beatus	Leyte	12	0.0070	0.0074
Crocidura beatus	Samar	11	0.0101	0.0097
Crocidura negrina-panayensis	Negros	12	0.0034	0.0051
Crocidura negrina-panayensis	Panay	6	0.0026	0.0020
Cynopterus brachyotis	Biliran	8	0.0096	0.0076
Cynopterus brachyotis	Mindanao	20	0.0108	0.0079
Cynopterus brachyotis	Negros	8	0.0032	0.0024
Cynopterus brachyotis	Panay	14	0.0121	0.0065
Cyrtodactylus annulatus	Bohol	3	0.0042	0.0045
Cyrtodactylus annulatus	Mindanao	14	0.0219	0.0239
Cyrtodactylus gubaot-sumuroi	Leyte	29	0.0129	0.0149
Cyrtodactylus gubaot-sumuroi	Samar	6	0.0046	0.0037
Cyrtodactylus philippinicus	Negros	6	0.0321	0.0381
Cyrtodactylus philippinicus	Panay	14	0.0202	0.0209
Dendrelaphis marenae	Negros	6	0.0008	0.0008
Dendrelaphis marenae	Panay	6	0.0004	0.0003
Gekko mindorensis	Negros	8	0.0233	0.0234
Gekko mindorensis	Panay	11	0.0148	0.0143
Haplonycteris fischeri	Biliran	8	0.0021	0.0025
Haplonycteris fischeri	Mindanao	29	0.0135	0.0076
Haplonycteris fischeri	Negros	9	0.0104	0.0081
Haplonycteris fischeri	Panay	21	0.0094	0.0057
Hipposideros obscurus	Leyte	9	0.0053	0.0044
Hipposideros obscurus	Mindanao	19	0.0063	0.0064
Hipposideros pygmaeus	Bohol	12	0.0019	0.0024
Hipposideros pygmaeus	Mindanao	3	0.0083	0.0083
Insulasaurus arborens	Negros	22	0.0042	0.0042
Insulasaurus arborens	Panay	10	0.0270	0.0310
Limnonectes leytensis	Bohol	2	0.0066	0.0066
Limnonectes leytensis	Mindanao	4	0.0114	0.0108
Limnonectes magnus	Bohol	3	0.0016	0.0016
Limnonectes magnus	Mindanao	2	0.0008	0.0008
Macroglossus minimus	Biliran	4	0.0076	0.0072
Macroglossus minimus	Mindanao	19	0.0109	0.0057
Macroglossus minimus	Negros	8	0.0161	0.0144
Macroglossus minimus	Panay	10	0.0153	0.0123
Pinoyscincus jagori	Mindanao	8	0.0011	0.0017
Pinoyscincus jagori	Samar	8	0.0045	0.0045
Ptenochirus jagori	Leyte	7	0.0138	0.0108
Ptenochirus jagori	Mindanao	4	0.0096	0.0088
Ptenochirus jagori	Negros	8	0.0091	0.0085
Ptenochirus jagori	Panay	8	0.0096	0.0085
Ptenochirus minor	Biliran	9	0.0049	0.0050
Ptenochirus minor	Mindanao	30	0.0107	0.0071

Table S5.6. Summaries of the posterior estimates from all `msBayes` analyses run with prior settings of $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$. Mode estimates of the dispersion index (Ω) and mean ($E(\tau)$) of divergence time vector, $\boldsymbol{\tau}$, and the number of unique divergence times (Ψ) are provided, followed in square brackets by the 95% highest posterior density for ABC_{GLM} analyses and the 2.5% and 97.5% quantiles for the ABC_{LLR} analyses. The estimated posterior probability for the one divergence model ($p(\Psi = 1|B_\epsilon(\mathbf{S}^*))$) is given, followed in parentheses by the Bayes factor of the $\Psi = 1$ model compared to all other models. Ω and $E(\tau)$ are in units of N_C generations.

Prior sample	**Analysis**	$\hat{\Omega}$	$E(\hat{\tau})$	$\hat{\Psi}$	$p(\Psi = 1\|B_\epsilon(\mathbf{S}^*))$
$\mathbb{S}_{stats,N=1\times10^7}$	ABC_{LLR}	3.49×10^{-4} [0–1.35×10^{-2}]	0.044 [0.015–0.080]	1.0 [1.00–1.00]	1.000 (∞)[1]
$\mathbb{S}_{stats,N=1\times10^7}$	ABC_{GLM}	2.64×10^{-3} [-1.01×10^{-5}–1.21×10^{-2}]	0.043 [0.014–0.074]	1.0 [1.00–1.06]	1.000 (∞)[1]
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	-9.35×10^{-17} [-1.01×10^{-5}–1.02×10^{-2}]	0.045 [0.012–0.078]	1.0 [1.00–1.06]	1.000 (∞)[1]
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	2.77×10^{-3} [-1.34×10^{-5}–1.54×10^{-2}]	0.040 [0.007–0.074]	1.0 [1.00–1.12]	0.998 (1.0×10^4)
$\mathbb{S}_{PLS,N=1\times10^7}$	ABC_{LLR}	0.0 [0.0–0.0]	0.055 [0–0.116]	1.0 [1.00–1.00]	0.999 (2.1×10^4)
$\mathbb{S}_{PLS,N=1\times10^7}$	ABC_{GLM}	-9.35×10^{-17} [-9.73×10^{-6}–9.95×10^{-3}]	0.034 [3.06×10^{-4}–0.095]	1.0 [1.00–1.13]	0.989 (1888)
$\mathbb{S}_{PLS,N=5\times10^6}$	ABC_{GLM}	-9.35×10^{-17} [-1.17×10^{-5}–1.39×10^{-2}]	0.031 [5.32×10^{-5}–0.089]	1.0 [1.00–1.12]	0.999 (2.1×10^4)
$\mathbb{S}_{PLS,N=5\times10^6}$	ABC_{GLM}	-8.86×10^{-17} [-8.14×10^{-6}–1.19×10^{-2}]	0.036 [2.00×10^{-4}–0.102]	1.0 [1.00–1.17]	0.926 (263)
13 Mindanao population-pairs					
$\mathbb{S}_{stats,N=2\times10^6}$	ABC_{LLR}	0.0 [0–2.03×10^{-2}]	0.056 [0.004–0.123]	1.0 [1.00–1.00]	0.999 (2.1×10^4)
$\mathbb{S}_{stats,N=2\times10^6}$	ABC_{GLM}	4.07×10^{-3} [-2.55×10^{-4}–3.07×10^{-2}]	0.053 [0.008–0.105]	1.0 [1.00–1.12]	0.999 (2.1×10^4)
9 Negros-Panay population-pairs					
$\mathbb{S}_{stats,N=2\times10^6}$	ABC_{LLR}	1.88×10^{-4} [0–1.52×10^{-2}]	0.038 [0–0.101]	1.0 [1.00–1.00]	0.997 (6979)
$\mathbb{S}_{stats,N=2\times10^6}$	ABC_{GLM}	-9.36×10^{-17} [-2.32×10^{-4}–2.26×10^{-2}]	0.032 [6.96×10^{-4}–0.083]	1.0 [1.00–1.06]	0.996 (5229)

[1] An estimate of 1.0 for a posterior probability is an artifact of sampling error.

Table S5.7. Posterior estimates of `msBayes` from the empirical data under three different model priors, before and after the θ_A error in the source code was corrected. All results are from analyses using $\mathbb{S}_{stats}$ and ABC_{GLM}. Analyses with broad priors used 10^7 samples from the prior, whereas the analyses with the informed prior settings used 5×10^6 samples. Mode posterior estimates of Ω, $E(\tau)$, and Ψ are provided, followed in square brackets by the 95% highest posterior density. The estimated posterior probability for the inferred divergence model ($p(\Psi = \hat{\Psi}|B_\epsilon(\mathbf{S}^*))$) is given, followed in parentheses by the Bayes factor of the inferred model compared to all other models. The estimated posterior probability that Ψ is not one ($p(\Psi \neq 1|B_\epsilon(\mathbf{S}^*))$) is also given. Ω and $E(\tau)$ are in units of N_C generations.

θ_A bug	$\hat{\Omega}$	$\hat{E(\tau)}$	$\hat{\Psi}$	$p(\Psi = \hat{\Psi}\|B_\epsilon(\mathbf{S}^*))$	$p(\Psi \neq 1\|B_\epsilon(\mathbf{S}^*))$
$f(\Theta)$: $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$					
Source	2.64×10^{-3} [-1.01×10^{-5}–1.21×10^{-2}]	0.043 [0.014–0.074]	1.00 [1.00–1.06]	1.000 (∞)[1]	0.000
Corrected	1.29×10^{-4} [-1.08×10^{-6}–1.18×10^{-3}]	0.063 [0.039–0.092]	1.00 [1.00–1.06]	1.000 (∞)[1]	0.000
$f(\Theta)$: $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$					
Source	0.490 [0.142–0.647]	0.404 [0.264–0.508]	2.01 [1.81–3.08]	0.855 (124)	0.994
Corrected	0.249 [0.030–0.475]	0.399 [0.271–0.513]	2.01 [1.39–3.40]	0.785 (77)	0.970
$f(\Theta)$: $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$					
Source	0.645 [0.255–0.852]	0.454 [0.264–0.541]	2.00 [1.55–4.09]	0.647 (38)	0.987
Corrected	0.441 [0.190–0.711]	0.513 [0.358–0.601]	2.04 [1.51–5.16]	0.343 (11)	0.986

[1] An estimate of 1.0 for a posterior probability is an artifact of sampling error.

Table S5.8. List of all data used in Chapter 1.

Genbank	Voucher ID	Species	Island	Gene	Use
AY313695	TNHC 56369	*Limnonectes leytensis*	Bohol	12S/16S	A,B
JX911318	RMB 2821[1]	*Limnonectes leytensis*	Bohol	12S/16S	A,B
JX911319	USNM 222545	*Limnonectes leytensis*	Leyte	12S/16S	B
AY313699	TNHC 59865	*Limnonectes leytensis*	Mindanao	12S/16S	A,B
AY313701	TNHC 59867	*Limnonectes leytensis*	Mindanao	12S/16S	A,B
JX911321	RMB 3777[1]	*Limnonectes leytensis*	Mindanao	12S/16S	A,B
JX911322	RMB 3788[1]	*Limnonectes leytensis*	Mindanao	12S/16S	A,B
JX911324	DEPOSITED IN PNM	*Limnonectes magnus*	Bohol	12S/16S	A,B
JX911326	PNM 7607	*Limnonectes magnus*	Bohol	12S/16S	A,B
JX911325	TNHC 56398	*Limnonectes magnus*	Bohol	12S/16S	A,B
AY313705	PNM 7449	*Limnonectes magnus*	Mindanao	12S/16S	A,B
JX911323	ACD 937[2]	*Limnonectes magnus*	Mindanao	12S/16S	A,B
JX911320	USNM 222570	*Limnonectes magnus*	Samar	12S/16S	B
AY313707	PNM 7600	*Limnonectes woodworthi*	Luzon	12S/16S	C
AY313709	TNHC 61919	*Limnonectes woodworthi*	Luzon	12S/16S	C
AY313710	TNHC 61922	*Limnonectes woodworthi*	Luzon	12S/16S	C
AY313711	TNHC 61916	*Limnonectes woodworthi*	Luzon	12S/16S	C
JX911331	RMB 4092[1]	*Limnonectes woodworthi*	Luzon	12S/16S	C
JX911327	ACD 949[2]	*Limnonectes woodworthi*	Luzon	12S/16S	C
JX911329	TNHC 61915	*Limnonectes woodworthi*	Luzon	12S/16S	C
JX911328	TNHC 61918	*Limnonectes woodworthi*	Luzon	12S/16S	C
JX911330	TNHC 61920	*Limnonectes woodworthi*	Luzon	12S/16S	C
JX678832	KU 305578	*Dendrelaphis caudolineatus*	Cebu	CYTB	B
JX678840	KU 310379	*Dendrelaphis caudolineatus*	Dinagat	CYTB	B
JX678808	ACD 1504[2]	*Dendrelaphis caudolineatus*	Leyte	CYTB	B
JX678821	KU 311282	*Dendrelaphis caudolineatus*	Leyte	CYTB	B
JX678831	CDSGS 47[6]	*Dendrelaphis caudolineatus*	Masbate	CYTB	B
JX678816	KU 302998	*Dendrelaphis caudolineatus*	Panay	CYTB	B
JX678820	KU 310856	*Dendrelaphis caudolineatus*	Samar	CYTB	B
JX678825	KU 303001	*Dendrelaphis marenae*	Guimaras	CYTB	B,C
JX678826	KU 303002	*Dendrelaphis marenae*	Guimaras	CYTB	B,C
JX678827	KU 303003	*Dendrelaphis marenae*	Guimaras	CYTB	B,C
JX678828	KU 303004	*Dendrelaphis marenae*	Guimaras	CYTB	B,C
JX678829	KU 303005	*Dendrelaphis marenae*	Guimaras	CYTB	B,C
JX678830	KU 303006	*Dendrelaphis marenae*	Guimaras	CYTB	B,C
JX678819	KU 305169	*Dendrelaphis marenae*	Luzon	CYTB	B,C
JX678833	KU 305581	*Dendrelaphis marenae*	Luzon	CYTB	B,C
JX678834	KU 305582	*Dendrelaphis marenae*	Luzon	CYTB	B,C
JX678835	KU 305583	*Dendrelaphis marenae*	Luzon	CYTB	B,C
JX678836	KU 305584	*Dendrelaphis marenae*	Luzon	CYTB	B,C
JX678837	RMB 3598[1]	*Dendrelaphis marenae*	Luzon	CYTB	B,C
JX678809	KU 303020	*Dendrelaphis marenae*	Negros	CYTB	A,B
JX678817	KU 305585	*Dendrelaphis marenae*	Negros	CYTB	A,B
JX678818	KU 305586	*Dendrelaphis marenae*	Negros	CYTB	A,B
JX678822	KU 303021	*Dendrelaphis marenae*	Negros	CYTB	A,B
JX678823	KU 303022	*Dendrelaphis marenae*	Negros	CYTB	A,B
JX678824	KU 303023	*Dendrelaphis marenae*	Negros	CYTB	A,B
JX678810	KU 303013	*Dendrelaphis marenae*	Panay	CYTB	A,B
JX678811	KU 303014	*Dendrelaphis marenae*	Panay	CYTB	A,B
JX678812	KU 303015	*Dendrelaphis marenae*	Panay	CYTB	A,B
JX678813	KU 303016	*Dendrelaphis marenae*	Panay	CYTB	A,B
JX678814	KU 303017	*Dendrelaphis marenae*	Panay	CYTB	A,B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
JX678815	KU 303019	*Dendrelaphis marenae*	Panay	CYTB	A,B
JX678838	KU 304103	*Dendrelaphis marenae*	Polillo	CYTB	B
JX678839	KU 307456	*Dendrelaphis marenae*	Polillo	CYTB	B
GU550762	KU 326502	*Cyrtodactylus annulatus*	Bohol	ND2	A,B
JX678802	TNHC 56467	*Cyrtodactylus annulatus*	Bohol	ND2	A,B
JX678803	KU 326501	*Cyrtodactylus annulatus*	Bohol	ND2	A,B
GU550763	KU 309366	*Cyrtodactylus annulatus*	Camiguin Sur	ND2	B
GU550764	KU 309364	*Cyrtodactylus annulatus*	Camiguin Sur	ND2	B
JX678804	KU 309365	*Cyrtodactylus annulatus*	Camiguin Sur	ND2	B
JX678805	KU 309363	*Cyrtodactylus annulatus*	Camiguin Sur	ND2	B
JX678795	PNM 1417	*Cyrtodactylus annulatus*	Leyte	ND2	B
JX678798	KU 311157	*Cyrtodactylus annulatus*	Leyte	ND2	B
GU550753	PNM 1473	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
GU550754	PNM 1470	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
GU550755	PNM/CMNH-H 1602[5]	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
GU550756	PNM/CMNH-H 1459[5]	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
GU550757	PNM/CMNH-H 1544[5]	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
GU550758	PNM/CMNH-H 1639[5]	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
GU550759	PNM/CMNH-H 1601[5]	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
JX678796	PNM 1472	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
JX678797	PNM 1471	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
JX678799	EMD 258[4]	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
JX678800	PNM/CMNH-H 1646[5]	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
JX678801	PNM/CMNH-H 1503[5]	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
JX678806	KU 314944	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
JX678807	KU 314945	*Cyrtodactylus annulatus*	Mindanao	ND2	A,B
GU550777	PNM 1418	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550778	PNM 1414	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550779	KU 309336	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550780	KU 311140	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550781	KU 311131	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550782	KU 311132	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550783	KU 311139	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550784	KU 309339	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550785	KU 311142	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550786	KU 311154	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550787	KU 311156	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550788	KU 311167	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550789	KU 311184	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550790	KU 311202	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550791	KU 311203	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550792	KU 311206	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550793	KU 311211	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550794	KU 311212	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550795	KU 311213	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550796	KU 311195	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550797	KU 311155	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550798	KU 311170	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550799	KU 311171	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550800	KU 311172	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550801	KU 311196	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550802	KU 309346	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550803	KU 311133	*Cyrtodactylus gubaot*	Leyte	ND2	A,B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
GU550804	KU 311134	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550805	KU 311137	*Cyrtodactylus gubaot*	Leyte	ND2	A,B
GU550806	KU 310094	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550807	KU 310096	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550808	KU 310097	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550809	KU 310105	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550810	KU 310106	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550811	KU 310107	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550812	KU 310109	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550813	KU 310110	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550814	KU 310111	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550815	KU 310112	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550816	KU 310098	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550817	KU 310099	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550818	KU 310100	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550819	KU 310101	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550820	KU 310102	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550821	KU 310103	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550822	KU 310104	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550823	KU 305564	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550824	KU 305565	*Cyrtodactylus mamanwa*	Dinagat	ND2	B,C
GU550902	KU 304587	*Cyrtodactylus philippinicus*	Camiguin Norte	ND2	B,C
GU550903	KU 304671	*Cyrtodactylus philippinicus*	Camiguin Norte	ND2	B,C
GU550904	KU 304631	*Cyrtodactylus philippinicus*	Camiguin Norte	ND2	B,C
GU550905	KU 304632	*Cyrtodactylus philippinicus*	Camiguin Norte	ND2	B,C
GU550906	KU 304634	*Cyrtodactylus philippinicus*	Camiguin Norte	ND2	B,C
GU550907	KU 304604	*Cyrtodactylus philippinicus*	Camiguin Norte	ND2	B,C
GU550908	KU 304635	*Cyrtodactylus philippinicus*	Camiguin Norte	ND2	B,C
GU550909	KU 304690	*Cyrtodactylus philippinicus*	Camiguin Norte	ND2	B,C
GU550910	KU 307963	*Cyrtodactylus philippinicus*	Camiguin Norte	ND2	B,C
GU550911	KU 308044	*Cyrtodactylus philippinicus*	Camiguin Norte	ND2	B,C
GU550912	KU 308140	*Cyrtodactylus philippinicus*	Catanduanes	ND2	B
GU550913	KU 308141	*Cyrtodactylus philippinicus*	Catanduanes	ND2	B
GU550872	RMB 942[1]	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550873	RMB 943[1]	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550874	KU 327075	*Cyrtodactylus philippinicus*	Luzon	ND2	B,C
GU550875	ACD 1976[2]	*Cyrtodactylus philippinicus*	Luzon	ND2	B,C
GU550876	KU 308256	*Cyrtodactylus philippinicus*	Luzon	ND2	B,C
GU550877	KU 327076	*Cyrtodactylus philippinicus*	Luzon	ND2	B,C
GU550878	KU 327072	*Cyrtodactylus philippinicus*	Luzon	ND2	B,C
GU550879	KU 327073	*Cyrtodactylus philippinicus*	Luzon	ND2	B,C
GU550880	RMB 4094[1]	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550881	KU 327074	*Cyrtodactylus philippinicus*	Luzon	ND2	B,C
GU550882	ACD 2094[2]	*Cyrtodactylus philippinicus*	Luzon	ND2	B,C
GU550883	KU 327071	*Cyrtodactylus philippinicus*	Luzon	ND2	B,C
GU550884	KU 327077	*Cyrtodactylus philippinicus*	Luzon	ND2	B,C
GU550886	TNHC 62840	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550887	TNHC 62841	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550888	TNHC 62835	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550889	RMB 3843[1]	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550890	TNHC 62839	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550891	RMB 3384[1]	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550892	TNHC 62837	*Cyrtodactylus philippinicus*	Luzon	ND2	B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
GU550893	TNHC 62836	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550894	KU 326527	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550895	KU 326528	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550896	TNHC 62838	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550897	KU 308814	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550898	KU 308815	*Cyrtodactylus philippinicus*	Luzon	ND2	B
GU550866	TNHC 62833	*Cyrtodactylus philippinicus*	Negros	ND2	A,B
GU550867	KU 305572	*Cyrtodactylus philippinicus*	Negros	ND2	A,B
GU550868	TNHC 62832	*Cyrtodactylus philippinicus*	Negros	ND2	A,B
GU550869	KU 302634	*Cyrtodactylus philippinicus*	Negros	ND2	A,B
GU550870	CDSGS 42[6]	*Cyrtodactylus philippinicus*	Negros	ND2	A,B
GU550871	CDSGS 43[6]	*Cyrtodactylus philippinicus*	Negros	ND2	A,B
GU550852	KU 302638	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550853	KU 302637	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550854	KU 302640	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550855	KU 302643	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550856	KU 302636	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550857	KU 302645	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550858	KU 302639	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550859	KU 302644	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550860	KU 302642	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550861	TNHC 56339	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550862	KU 306612	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550863	GVAG 257[3]	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550864	GVAG 226[3]	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550865	KU 306613	*Cyrtodactylus philippinicus*	Panay	ND2	A,B
GU550914	KU 303827	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550915	KU 307453	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550916	KU 307455	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550917	KU 302647	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550918	KU 307451	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550919	KU 302650	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550920	KU 303826	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550921	KU 302646	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550922	KU 307452	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550923	KU 303899	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550924	KU 303825	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550925	KU 302648	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550926	KU 302649	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550927	KU 307454	*Cyrtodactylus philippinicus*	Polillo	ND2	B,C
GU550928	KU 303852	*Cyrtodactylus philippinicus*	Romblon	ND2	B
GU550929	RMB 2962[1]	*Cyrtodactylus philippinicus*	Sibuyan	ND2	B
GU550930	KU 303849	*Cyrtodactylus philippinicus*	Sibuyan	ND2	B
GU550931	KU 302635	*Cyrtodactylus philippinicus*	Sibuyan	ND2	B
GU550932	KU 303848	*Cyrtodactylus philippinicus*	Sibuyan	ND2	B
GU550938	KU 315340	*Cyrtodactylus philippinicus*	Tablas	ND2	B
GU550939	KU 315341	*Cyrtodactylus philippinicus*	Tablas	ND2	B
GU550940	KU 315342	*Cyrtodactylus philippinicus*	Tablas	ND2	B
GU550941	KU 315343	*Cyrtodactylus philippinicus*	Tablas	ND2	B
GU550942	KU 315344	*Cyrtodactylus philippinicus*	Tablas	ND2	B
GU550885	KU 327078	*Cyrtodactylus philippinicus*		ND2	B
GU550771	KU 305566	*Cyrtodactylus sumuroi*	Samar	ND2	A,B
GU550772	KU 310798	*Cyrtodactylus sumuroi*	Samar	ND2	A,B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
GU550773	KU 310799	*Cyrtodactylus sumuroi*	Samar	ND2	A,B
GU550774	KU 310800	*Cyrtodactylus sumuroi*	Samar	ND2	A,B
GU550775	KU 310801	*Cyrtodactylus sumuroi*	Samar	ND2	A,B
GU550776	KU 310802	*Cyrtodactylus sumuroi*	Samar	ND2	A,B
JX678786	RMB 2245[1]	*Gekko gigante*	Gigante	ND2	B,C
JX678787	RMB 2246[1]	*Gekko gigante*	Gigante	ND2	B,C
JX678788	RMB 2247[1]	*Gekko gigante*	Gigante	ND2	B,C
JX678789	RMB 2248[1]	*Gekko gigante*	Gigante	ND2	B,C
JQ173417	KU 302716	*Gekko gigante*	N. Gigante	ND2	B,C
JQ173419	KU 302720	*Gekko gigante*	N. Gigante	ND2	B,C
JQ173420	KU 305138	*Gekko gigante*	N. Gigante	ND2	B,C
JQ173421	KU 305140	*Gekko gigante*	N. Gigante	ND2	B,C
JX678784	KU 305139	*Gekko gigante*	N. Gigante	ND2	B,C
JQ173418	KU 302718	*Gekko gigante*	S. Gigante	ND2	B,C
JX678785	KU 302719	*Gekko gigante*	S. Gigante	ND2	B,C
JQ173458	KU 302721	*Gekko mindorensis*	Guimaras	ND2	B
JQ173462	KU 302725	*Gekko mindorensis*	Guimaras	ND2	B
JQ173470	KU 302733	*Gekko mindorensis*	Maestre De Campo	ND2	B
JQ173471	KU 302734	*Gekko mindorensis*	Maestre De Campo	ND2	B
JQ173472	KU 302735	*Gekko mindorensis*	Maestre De Campo	ND2	B
JQ173463	KU 302726	*Gekko mindorensis*	Masbate	ND2	B
JQ173464	KU 302727	*Gekko mindorensis*	Masbate	ND2	B
JQ173465	KU 302728	*Gekko mindorensis*	Masbate	ND2	B
JQ173439	CDSGS 38[6]	*Gekko mindorensis*	Negros	ND2	A,B
JQ173440	CDSGS 39[6]	*Gekko mindorensis*	Negros	ND2	A,B
JQ173441	CDSGS 40[6]	*Gekko mindorensis*	Negros	ND2	A,B
JQ173442	CDSGS 41[6]	*Gekko mindorensis*	Negros	ND2	A,B
JQ173459	KU 302722	*Gekko mindorensis*	Negros	ND2	A,B
JQ173460	KU 302723	*Gekko mindorensis*	Negros	ND2	A,B
JQ173461	KU 302724	*Gekko mindorensis*	Negros	ND2	A,B
JX678794	RMB 3252[1]	*Gekko mindorensis*	Negros	ND2	A,B
JQ173437	CDSGS 03[6]	*Gekko mindorensis*	Panay	ND2	A,B
JQ173438	CDSGS 04[6]	*Gekko mindorensis*	Panay	ND2	A,B
JQ173443	GVAG 260[3]	*Gekko mindorensis*	Panay	ND2	A,B
JQ173466	KU 302729	*Gekko mindorensis*	Panay	ND2	A,B
JQ173467	KU 302730	*Gekko mindorensis*	Panay	ND2	A,B
JQ173468	KU 302731	*Gekko mindorensis*	Panay	ND2	A,B
JQ173469	KU 302732	*Gekko mindorensis*	Panay	ND2	A,B
JQ173491	KU 307060	*Gekko mindorensis*	Panay	ND2	A,B
JQ173492	KU 307061	*Gekko mindorensis*	Panay	ND2	A,B
JX678790	KU 307058	*Gekko mindorensis*	Panay	ND2	A,B
JX678791	KU 307059	*Gekko mindorensis*	Panay	ND2	A,B
JX678792	KU 307063	*Gekko mindorensis*	Panay	ND2	B
JX678793	KU 307065	*Gekko mindorensis*	Panay	ND2	B
JF498114	KU 306712	*Insulasaurus arborens*	Negros	ND2	A,B
JX943478	KU 306700	*Insulasaurus arborens*	Negros	ND2	A,B
JX943479	KU 306701	*Insulasaurus arborens*	Negros	ND2	A,B
JX943480	KU 306702	*Insulasaurus arborens*	Negros	ND2	A,B
JX943481	KU 306703	*Insulasaurus arborens*	Negros	ND2	A,B
JX943482	KU 306704	*Insulasaurus arborens*	Negros	ND2	A,B
JX943483	KU 306705	*Insulasaurus arborens*	Negros	ND2	A,B
JX943484	KU 306706	*Insulasaurus arborens*	Negros	ND2	A,B
JX943485	KU 306707	*Insulasaurus arborens*	Negros	ND2	A,B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
JX943486	KU 306709	*Insulasaurus arborens*	Negros	ND2	A,B
JX943487	KU 306710	*Insulasaurus arborens*	Negros	ND2	A,B
JX943488	KU 306711	*Insulasaurus arborens*	Negros	ND2	A,B
JX943489	KU 306713	*Insulasaurus arborens*	Negros	ND2	A,B
JX943490	KU 306714	*Insulasaurus arborens*	Negros	ND2	A,B
JX943491	KU 306715	*Insulasaurus arborens*	Negros	ND2	A,B
JX943492	KU 306716	*Insulasaurus arborens*	Negros	ND2	A,B
JX943493	KU 306717	*Insulasaurus arborens*	Negros	ND2	A,B
JX943494	KU 306718	*Insulasaurus arborens*	Negros	ND2	A,B
JX943495	KU 306719	*Insulasaurus arborens*	Negros	ND2	A,B
JX943496	KU 306720	*Insulasaurus arborens*	Negros	ND2	A,B
JX943498	TNHC 62465	*Insulasaurus arborens*	Negros	ND2	A,B
JX943497	RMB 7248[1]	*Insulasaurus arborens*	Negros	ND2	A,B
JF498113	KU 306805	*Insulasaurus arborens*	Panay	ND2	A,B
JX943505	KU 306806	*Insulasaurus arborens*	Panay	ND2	A,B
JX943507	KU 306807	*Insulasaurus arborens*	Panay	ND2	A,B
JX943506	KU 306830	*Insulasaurus arborens*	Panay	ND2	A,B
JX943500	PNM 572	*Insulasaurus arborens*	Panay	ND2	A,B
JX943501	PNM 575	*Insulasaurus arborens*	Panay	ND2	A,B
JX943502	PNM 586	*Insulasaurus arborens*	Panay	ND2	A,B
JX943503	PNM 588	*Insulasaurus arborens*	Panay	ND2	A,B
JX943504	PNM 609	*Insulasaurus arborens*	Panay	ND2	A,B
JX943499	GVAG 258[3]	*Insulasaurus arborens*	Panay	ND2	A,B
GU573648	TNHC 62758	*Pinoyscincus abdictus*	Luzon	ND2	B
GU573649	KU 307679	*Pinoyscincus abdictus*	Polillo	ND2	B,C
GU573650	KU 307688	*Pinoyscincus abdictus*	Polillo	ND2	B,C
GU573651	KU 302921	*Pinoyscincus abdictus*	Polillo	ND2	B,C
GU573652	KU 302916	*Pinoyscincus abdictus*	Polillo	ND2	B,C
GU573653	KU 302912	*Pinoyscincus abdictus*	Polillo	ND2	B,C
GU573654	KU 302917	*Pinoyscincus abdictus*	Polillo	ND2	B,C
GU573655	KU 302911	*Pinoyscincus abdictus*	Polillo	ND2	B,C
GU573656	KU 302919	*Pinoyscincus abdictus*	Polillo	ND2	B,C
GU573657	KU 302914	*Pinoyscincus abdictus*	Polillo	ND2	B,C
GU573658	KU 307689	*Pinoyscincus abdictus*	Polillo	ND2	B,C
GU573659	KU 307680	*Pinoyscincus abdictus*	Polillo	ND2	B,C
GU573661	KU 302915	*Pinoyscincus abdictus*	Polillo	ND2	B,C
GU573585	TNHC 56380	*Pinoyscincus jagori*	Bohol	ND2	B
GU573586	KU 326402	*Pinoyscincus jagori*	Bohol	ND2	B
GU573587	KU 306546	*Pinoyscincus jagori*	Dinagat	ND2	B
GU573588	KU 306547	*Pinoyscincus jagori*	Dinagat	ND2	B
GU573589	KU 315067	*Pinoyscincus jagori*	Mindanao	ND2	B
GU573590	KU 315069	*Pinoyscincus jagori*	Mindanao	ND2	B
JX943508	EMD 219[4]	*Pinoyscincus jagori*	Mindanao	ND2	A,B
JX943509	EMD 220[4]	*Pinoyscincus jagori*	Mindanao	ND2	A,B
JX943510	EMD 222[4]	*Pinoyscincus jagori*	Mindanao	ND2	A,B
JX943511	EMD 225[4]	*Pinoyscincus jagori*	Mindanao	ND2	A,B
JX943512	EMD 226[4]	*Pinoyscincus jagori*	Mindanao	ND2	A,B
JX943513	EMD 228[4]	*Pinoyscincus jagori*	Mindanao	ND2	A,B
JX943514	EMD 247[4]	*Pinoyscincus jagori*	Mindanao	ND2	A,B
JX943515	EMD 248[4]	*Pinoyscincus jagori*	Mindanao	ND2	A,B
GU573576	KU 306550	*Pinoyscincus jagori*	Samar	ND2	A,B
GU573577	KU 306555	*Pinoyscincus jagori*	Samar	ND2	A,B
GU573578	KU 306554	*Pinoyscincus jagori*	Samar	ND2	A,B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
GU573579	KU 306548	*Pinoyscincus jagori*	Samar	ND2	A,B
GU573580	KU 306553	*Pinoyscincus jagori*	Samar	ND2	A,B
GU573581	KU 306552	*Pinoyscincus jagori*	Samar	ND2	A,B
GU573582	KU 306551	*Pinoyscincus jagori*	Samar	ND2	A,B
GU573583	KU 306549	*Pinoyscincus jagori*	Samar	ND2	A,B
AY922376	[8]	*Cynopterus brachyotis*	Biliran	ND2	A,B
AY922377	[8]	*Cynopterus brachyotis*	Biliran	ND2	A,B
AY922378	[8]	*Cynopterus brachyotis*	Biliran	ND2	A,B
AY922379	[8]	*Cynopterus brachyotis*	Biliran	ND2	A,B
AY922380	[8]	*Cynopterus brachyotis*	Biliran	ND2	A,B
AY922381	[8]	*Cynopterus brachyotis*	Biliran	ND2	A,B
AY922382	[8]	*Cynopterus brachyotis*	Biliran	ND2	A,B
AY922383	[8]	*Cynopterus brachyotis*	Biliran	ND2	A,B
AY926607	[8]	*Cynopterus brachyotis*	Borneo	ND2	B
AY926608	[8]	*Cynopterus brachyotis*	Borneo	ND2	B
AY926609	[8]	*Cynopterus brachyotis*	Borneo	ND2	B
AY926610	[8]	*Cynopterus brachyotis*	Borneo	ND2	B
AY926611	[8]	*Cynopterus brachyotis*	Borneo	ND2	B
AY922384	[8]	*Cynopterus brachyotis*	Camiguin	ND2	B,C
AY922385	[8]	*Cynopterus brachyotis*	Camiguin	ND2	B,C
AY922386	[8]	*Cynopterus brachyotis*	Camiguin	ND2	B,C
AY922387	[8]	*Cynopterus brachyotis*	Camiguin	ND2	B,C
AY922388	[8]	*Cynopterus brachyotis*	Camiguin	ND2	B,C
AY922389	[8]	*Cynopterus brachyotis*	Camiguin	ND2	B,C
AY922390	[8]	*Cynopterus brachyotis*	Camiguin	ND2	B,C
AY922391	[8]	*Cynopterus brachyotis*	Camiguin	ND2	B,C
AY922392	[8]	*Cynopterus brachyotis*	Catanduanes	ND2	B
AY922393	[8]	*Cynopterus brachyotis*	Catanduanes	ND2	B
AY922394	[8]	*Cynopterus brachyotis*	Catanduanes	ND2	B
AY922395	[8]	*Cynopterus brachyotis*	Catanduanes	ND2	B
AY922396	[8]	*Cynopterus brachyotis*	Catanduanes	ND2	B
AY922397	[8]	*Cynopterus brachyotis*	Catanduanes	ND2	B
AY922398	[8]	*Cynopterus brachyotis*	Catanduanes	ND2	B
AY922430	[8]	*Cynopterus brachyotis*	Leyte	ND2	B
AY922431	[8]	*Cynopterus brachyotis*	Leyte	ND2	B
AY922432	[8]	*Cynopterus brachyotis*	Leyte	ND2	B
AY922433	[8]	*Cynopterus brachyotis*	Leyte	ND2	B
AY922434	[8]	*Cynopterus brachyotis*	Leyte	ND2	B
AY922435	[8]	*Cynopterus brachyotis*	Leyte	ND2	B
AY922414	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922415	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922416	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922417	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922418	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922419	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922420	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922421	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922422	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922423	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922424	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922425	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922426	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922427	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
AY922428	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922429	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922485	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922486	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922487	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922488	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922489	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922490	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922491	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922492	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922493	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922494	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922495	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922496	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922497	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY922498	[8]	*Cynopterus brachyotis*	Luzon	ND2	B,C
AY926612	[8]	*Cynopterus brachyotis*	Malaya	ND2	B
AY926613	[8]	*Cynopterus brachyotis*	Malaya	ND2	B
AY926619	[8]	*Cynopterus brachyotis*	Malaya	ND2	B
AY922399	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922400	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922401	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922402	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922403	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922404	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922405	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922406	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922407	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922436	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922437	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922438	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922439	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922440	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922441	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922442	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922443	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922467	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922468	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922469	[8]	*Cynopterus brachyotis*	Mindanao	ND2	A,B
AY922444	[8]	*Cynopterus brachyotis*	Negros	ND2	A,B
AY922445	[8]	*Cynopterus brachyotis*	Negros	ND2	A,B
AY922446	[8]	*Cynopterus brachyotis*	Negros	ND2	A,B
AY922447	[8]	*Cynopterus brachyotis*	Negros	ND2	A,B
AY922448	[8]	*Cynopterus brachyotis*	Negros	ND2	A,B
AY922449	[8]	*Cynopterus brachyotis*	Negros	ND2	A,B
AY922450	[8]	*Cynopterus brachyotis*	Negros	ND2	A,B
AY922451	[8]	*Cynopterus brachyotis*	Negros	ND2	A,B
AY926614	[8]	*Cynopterus brachyotis*	Palawan	ND2	B
AY926615	[8]	*Cynopterus brachyotis*	Palawan	ND2	B
AY926616	[8]	*Cynopterus brachyotis*	Palawan	ND2	B
AY926617	[8]	*Cynopterus brachyotis*	Palawan	ND2	B
AY926618	[8]	*Cynopterus brachyotis*	Palawan	ND2	B
AY922408	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
AY922409	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922410	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922411	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922412	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922413	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922452	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922453	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922454	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922455	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922456	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922457	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922458	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922459	[8]	*Cynopterus brachyotis*	Panay	ND2	A,B
AY922460	[8]	*Cynopterus brachyotis*	Polillo	ND2	B,C
AY922461	[8]	*Cynopterus brachyotis*	Polillo	ND2	B,C
AY922462	[8]	*Cynopterus brachyotis*	Polillo	ND2	B,C
AY922463	[8]	*Cynopterus brachyotis*	Polillo	ND2	B,C
AY922464	[8]	*Cynopterus brachyotis*	Polillo	ND2	B,C
AY922465	[8]	*Cynopterus brachyotis*	Polillo	ND2	B,C
AY922466	[8]	*Cynopterus brachyotis*	Polillo	ND2	B,C
AY922470	[8]	*Cynopterus brachyotis*	Sibuyan	ND2	B,C
AY922471	[8]	*Cynopterus brachyotis*	Sibuyan	ND2	B,C
AY922472	[8]	*Cynopterus brachyotis*	Sibuyan	ND2	B,C
AY922473	[8]	*Cynopterus brachyotis*	Sibuyan	ND2	B,C
AY922474	[8]	*Cynopterus brachyotis*	Sibuyan	ND2	B,C
AY922475	[8]	*Cynopterus brachyotis*	Sibuyan	ND2	B,C
AY922476	[8]	*Cynopterus brachyotis*	Sibuyan	ND2	B,C
AY922477	[8]	*Cynopterus brachyotis*	Sibuyan	ND2	B,C
AY922478	[8]	*Cynopterus brachyotis*	Siquijor	ND2	B,C
AY922479	[8]	*Cynopterus brachyotis*	Siquijor	ND2	B,C
AY922480	[8]	*Cynopterus brachyotis*	Siquijor	ND2	B,C
AY922481	[8]	*Cynopterus brachyotis*	Siquijor	ND2	B,C
AY922482	[8]	*Cynopterus brachyotis*	Siquijor	ND2	B,C
AY922483	[8]	*Cynopterus brachyotis*	Siquijor	ND2	B,C
AY922484	[8]	*Cynopterus brachyotis*	Siquijor	ND2	B,C
AY926620	[8]	*Cynopterus brachyotis*	Sulawesi	ND2	B
AY926621	[8]	*Cynopterus brachyotis*	Sulawesi	ND2	B
AY926622	[8]	*Cynopterus brachyotis*	Sulawesi	ND2	B
AY817769	[8]	*Haplonycteris fischeri*	Biliran	ND2	A,B
AY817770	[8]	*Haplonycteris fischeri*	Biliran	ND2	A,B
AY817771	[8]	*Haplonycteris fischeri*	Biliran	ND2	A,B
AY817772	[8]	*Haplonycteris fischeri*	Biliran	ND2	A,B
AY817773	[8]	*Haplonycteris fischeri*	Biliran	ND2	A,B
AY817774	[8]	*Haplonycteris fischeri*	Biliran	ND2	A,B
AY817775	[8]	*Haplonycteris fischeri*	Biliran	ND2	A,B
AY817776	[8]	*Haplonycteris fischeri*	Biliran	ND2	A,B
AY817777	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C
AY817778	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C
AY817779	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C
AY817780	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C
AY817781	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C
AY817782	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C
AY817783	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
AY817784	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C
AY817785	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C
AY817786	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C
AY817787	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C
AY817788	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C
AY817789	[8]	*Haplonycteris fischeri*	Catanduanes	ND2	B,C
AY817804	[8]	*Haplonycteris fischeri*	Leyte	ND2	B
AY817877	[8]	*Haplonycteris fischeri*	Luzon	ND2	B
AY817878	[8]	*Haplonycteris fischeri*	Luzon	ND2	B
AY817879	[8]	*Haplonycteris fischeri*	Luzon	ND2	B
AY817880	[8]	*Haplonycteris fischeri*	Luzon	ND2	B
AY817790	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817791	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817792	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817793	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817794	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817795	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817796	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817797	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817798	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817799	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817800	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817801	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817802	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817803	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817805	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817806	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817807	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817808	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817809	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817810	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817811	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817812	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817813	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817814	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817815	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817816	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817817	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817818	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817819	[8]	*Haplonycteris fischeri*	Mindanao	ND2	A,B
AY817847	[8]	*Haplonycteris fischeri*	Mindanao	ND2	B
AY817848	[8]	*Haplonycteris fischeri*	Mindanao	ND2	B
AY817849	[8]	*Haplonycteris fischeri*	Mindanao	ND2	B
AY817850	[8]	*Haplonycteris fischeri*	Mindanao	ND2	B
AY817851	[8]	*Haplonycteris fischeri*	Mindanao	ND2	B
AY817852	[8]	*Haplonycteris fischeri*	Mindanao	ND2	B
AY817853	[8]	*Haplonycteris fischeri*	Mindanao	ND2	B
AY817854	[8]	*Haplonycteris fischeri*	Mindanao	ND2	B
AY817855	[8]	*Haplonycteris fischeri*	Mindanao	ND2	B
AY817856	[8]	*Haplonycteris fischeri*	Mindanao	ND2	B
AY817829	[8]	*Haplonycteris fischeri*	Mindoro	ND2	B,C
AY817830	[8]	*Haplonycteris fischeri*	Mindoro	ND2	B,C
AY817831	[8]	*Haplonycteris fischeri*	Mindoro	ND2	B,C

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
AY817832	[8]	*Haplonycteris fischeri*	Mindoro	ND2	B,C
AY817833	[8]	*Haplonycteris fischeri*	Mindoro	ND2	B,C
AY817834	[8]	*Haplonycteris fischeri*	Mindoro	ND2	B,C
AY817835	[8]	*Haplonycteris fischeri*	Mindoro	ND2	B,C
AY817836	[8]	*Haplonycteris fischeri*	Mindoro	ND2	B,C
AY817820	[8]	*Haplonycteris fischeri*	Negros	ND2	A,B
AY817821	[8]	*Haplonycteris fischeri*	Negros	ND2	A,B
AY817822	[8]	*Haplonycteris fischeri*	Negros	ND2	A,B
AY817823	[8]	*Haplonycteris fischeri*	Negros	ND2	A,B
AY817824	[8]	*Haplonycteris fischeri*	Negros	ND2	A,B
AY817825	[8]	*Haplonycteris fischeri*	Negros	ND2	A,B
AY817826	[8]	*Haplonycteris fischeri*	Negros	ND2	A,B
AY817827	[8]	*Haplonycteris fischeri*	Negros	ND2	A,B
AY817828	[8]	*Haplonycteris fischeri*	Negros	ND2	A,B
AY817758	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817759	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817760	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817761	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817762	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817763	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817764	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817765	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817766	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817767	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817768	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817837	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817838	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817839	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817840	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817841	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817842	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817843	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817844	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817845	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817846	[8]	*Haplonycteris fischeri*	Panay	ND2	A,B
AY817857	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817858	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817859	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817860	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817861	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817862	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817863	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817864	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817865	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817866	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817867	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817868	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817869	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817870	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817871	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817872	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817873	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817874	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
AY817875	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY817876	[8]	*Haplonycteris fischeri*	Sibuyan	ND2	B,C
AY922622	[8]	*Macroglossus minimus*	Biliran	ND2	A,B
AY922623	[8]	*Macroglossus minimus*	Biliran	ND2	A,B
AY922624	[8]	*Macroglossus minimus*	Biliran	ND2	A,B
AY922625	[8]	*Macroglossus minimus*	Biliran	ND2	A,B
AY926639	[8]	*Macroglossus minimus*	Borneo	ND2	B
AY926640	[8]	*Macroglossus minimus*	Borneo	ND2	B
AY926641	[8]	*Macroglossus minimus*	Borneo	ND2	B
AY926642	[8]	*Macroglossus minimus*	Borneo	ND2	B
AY922626	[8]	*Macroglossus minimus*	Camiguin	ND2	B,C
AY922627	[8]	*Macroglossus minimus*	Camiguin	ND2	B,C
AY922628	[8]	*Macroglossus minimus*	Camiguin	ND2	B,C
AY922629	[8]	*Macroglossus minimus*	Camiguin	ND2	B,C
AY922630	[8]	*Macroglossus minimus*	Camiguin	ND2	B,C
AY922631	[8]	*Macroglossus minimus*	Camiguin	ND2	B,C
AY922632	[8]	*Macroglossus minimus*	Camiguin	ND2	B,C
AY922633	[8]	*Macroglossus minimus*	Camiguin	ND2	B,C
AY922634	[8]	*Macroglossus minimus*	Catanduanes	ND2	B
AY922635	[8]	*Macroglossus minimus*	Catanduanes	ND2	B
AY922636	[8]	*Macroglossus minimus*	Catanduanes	ND2	B
AY922637	[8]	*Macroglossus minimus*	Catanduanes	ND2	B
AY922643	[8]	*Macroglossus minimus*	Leyte	ND2	B
AY922644	[8]	*Macroglossus minimus*	Leyte	ND2	B
AY922645	[8]	*Macroglossus minimus*	Leyte	ND2	B
AY922641	[8]	*Macroglossus minimus*	Luzon	ND2	B,C
AY922642	[8]	*Macroglossus minimus*	Luzon	ND2	B,C
AY922699	[8]	*Macroglossus minimus*	Luzon	ND2	B,C
AY922700	[8]	*Macroglossus minimus*	Luzon	ND2	B,C
AY922701	[8]	*Macroglossus minimus*	Luzon	ND2	B,C
AY922702	[8]	*Macroglossus minimus*	Luzon	ND2	B,C
AY922703	[8]	*Macroglossus minimus*	Luzon	ND2	B,C
AY922704	[8]	*Macroglossus minimus*	Luzon	ND2	B,C
AY922705	[8]	*Macroglossus minimus*	Luzon	ND2	B,C
AY922706	[8]	*Macroglossus minimus*	Luzon	ND2	B,C
AY922707	[8]	*Macroglossus minimus*	Luzon	ND2	B,C
AY922638	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922639	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922640	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922646	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922647	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922648	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922649	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922650	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922651	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922652	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922653	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922675	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922676	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922677	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922678	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922679	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922680	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
AY922681	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922682	[8]	*Macroglossus minimus*	Mindanao	ND2	A,B
AY922662	[8]	*Macroglossus minimus*	Mindoro	ND2	B
AY922663	[8]	*Macroglossus minimus*	Mindoro	ND2	B
AY922664	[8]	*Macroglossus minimus*	Mindoro	ND2	B
AY922654	[8]	*Macroglossus minimus*	Negros	ND2	A,B
AY922655	[8]	*Macroglossus minimus*	Negros	ND2	A,B
AY922656	[8]	*Macroglossus minimus*	Negros	ND2	A,B
AY922657	[8]	*Macroglossus minimus*	Negros	ND2	A,B
AY922658	[8]	*Macroglossus minimus*	Negros	ND2	A,B
AY922659	[8]	*Macroglossus minimus*	Negros	ND2	A,B
AY922660	[8]	*Macroglossus minimus*	Negros	ND2	A,B
AY922661	[8]	*Macroglossus minimus*	Negros	ND2	A,B
AY926643	[8]	*Macroglossus minimus*	Palawan	ND2	B
AY926644	[8]	*Macroglossus minimus*	Palawan	ND2	B
AY922665	[8]	*Macroglossus minimus*	Panay	ND2	A,B
AY922666	[8]	*Macroglossus minimus*	Panay	ND2	A,B
AY922667	[8]	*Macroglossus minimus*	Panay	ND2	A,B
AY922668	[8]	*Macroglossus minimus*	Panay	ND2	A,B
AY922669	[8]	*Macroglossus minimus*	Panay	ND2	A,B
AY922670	[8]	*Macroglossus minimus*	Panay	ND2	A,B
AY922671	[8]	*Macroglossus minimus*	Panay	ND2	A,B
AY922672	[8]	*Macroglossus minimus*	Panay	ND2	A,B
AY922673	[8]	*Macroglossus minimus*	Panay	ND2	A,B
AY922674	[8]	*Macroglossus minimus*	Panay	ND2	A,B
AY922683	[8]	*Macroglossus minimus*	Sibuyan	ND2	B,C
AY922684	[8]	*Macroglossus minimus*	Sibuyan	ND2	B,C
AY922685	[8]	*Macroglossus minimus*	Sibuyan	ND2	B,C
AY922686	[8]	*Macroglossus minimus*	Sibuyan	ND2	B,C
AY922687	[8]	*Macroglossus minimus*	Sibuyan	ND2	B,C
AY922688	[8]	*Macroglossus minimus*	Sibuyan	ND2	B,C
AY922689	[8]	*Macroglossus minimus*	Sibuyan	ND2	B,C
AY922690	[8]	*Macroglossus minimus*	Sibuyan	ND2	B,C
AY922691	[8]	*Macroglossus minimus*	Siquijor	ND2	B,C
AY922692	[8]	*Macroglossus minimus*	Siquijor	ND2	B,C
AY922693	[8]	*Macroglossus minimus*	Siquijor	ND2	B,C
AY922694	[8]	*Macroglossus minimus*	Siquijor	ND2	B,C
AY922695	[8]	*Macroglossus minimus*	Siquijor	ND2	B,C
AY922696	[8]	*Macroglossus minimus*	Siquijor	ND2	B,C
AY922697	[8]	*Macroglossus minimus*	Siquijor	ND2	B,C
AY922698	[8]	*Macroglossus minimus*	Siquijor	ND2	B,C
AY922884	[8]	*Ptenochirus jagori*	Biliran	ND2	B
AY922885	[8]	*Ptenochirus jagori*	Biliran	ND2	B
AY922886	[8]	*Ptenochirus jagori*	Biliran	ND2	B
AY922887	[8]	*Ptenochirus jagori*	Camiguin	ND2	B,C
AY922888	[8]	*Ptenochirus jagori*	Camiguin	ND2	B,C
AY922889	[8]	*Ptenochirus jagori*	Camiguin	ND2	B,C
AY922890	[8]	*Ptenochirus jagori*	Camiguin	ND2	B,C
AY922891	[8]	*Ptenochirus jagori*	Camiguin	ND2	B,C
AY922892	[8]	*Ptenochirus jagori*	Camiguin	ND2	B,C
AY922893	[8]	*Ptenochirus jagori*	Camiguin	ND2	B,C
AY922894	[8]	*Ptenochirus jagori*	Camiguin	ND2	B,C
AY922895	[8]	*Ptenochirus jagori*	Catanduanes	ND2	B,C

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
AY922896	[8]	*Ptenochirus jagori*	Catanduanes	ND2	B,C
AY922897	[8]	*Ptenochirus jagori*	Catanduanes	ND2	B,C
AY922898	[8]	*Ptenochirus jagori*	Catanduanes	ND2	B,C
AY922899	[8]	*Ptenochirus jagori*	Catanduanes	ND2	B,C
AY922900	[8]	*Ptenochirus jagori*	Catanduanes	ND2	B,C
AY922901	[8]	*Ptenochirus jagori*	Catanduanes	ND2	B,C
AY922902	[8]	*Ptenochirus jagori*	Catanduanes	ND2	B,C
AY922918	[8]	*Ptenochirus jagori*	Leyte	ND2	A,B
AY922919	[8]	*Ptenochirus jagori*	Leyte	ND2	A,B
AY922920	[8]	*Ptenochirus jagori*	Leyte	ND2	A,B
AY922921	[8]	*Ptenochirus jagori*	Leyte	ND2	A,B
AY922922	[8]	*Ptenochirus jagori*	Leyte	ND2	A,B
AY922923	[8]	*Ptenochirus jagori*	Leyte	ND2	A,B
AY922924	[8]	*Ptenochirus jagori*	Leyte	ND2	A,B
AY922905	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922906	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922907	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922908	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922909	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922910	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922911	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922912	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922913	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922914	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922915	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922916	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922917	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922961	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922962	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922963	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922964	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922965	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922966	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922967	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922968	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922969	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922970	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922971	[8]	*Ptenochirus jagori*	Luzon	ND2	B,C
AY922925	[8]	*Ptenochirus jagori*	Mindanao	ND2	A,B
AY922926	[8]	*Ptenochirus jagori*	Mindanao	ND2	A,B
AY922927	[8]	*Ptenochirus jagori*	Mindanao	ND2	A,B
AY922928	[8]	*Ptenochirus jagori*	Mindanao	ND2	A,B
AY922929	[8]	*Ptenochirus jagori*	Negros	ND2	A,B
AY922930	[8]	*Ptenochirus jagori*	Negros	ND2	A,B
AY922931	[8]	*Ptenochirus jagori*	Negros	ND2	A,B
AY922932	[8]	*Ptenochirus jagori*	Negros	ND2	A,B
AY922933	[8]	*Ptenochirus jagori*	Negros	ND2	A,B
AY922934	[8]	*Ptenochirus jagori*	Negros	ND2	A,B
AY922935	[8]	*Ptenochirus jagori*	Negros	ND2	A,B
AY922936	[8]	*Ptenochirus jagori*	Negros	ND2	A,B
AY922883	[8]	*Ptenochirus jagori*	Panay	ND2	A,B
AY922903	[8]	*Ptenochirus jagori*	Panay	ND2	A,B
AY922904	[8]	*Ptenochirus jagori*	Panay	ND2	A,B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
AY922937	[8]	*Ptenochirus jagori*	Panay	ND2	A,B
AY922938	[8]	*Ptenochirus jagori*	Panay	ND2	A,B
AY922939	[8]	*Ptenochirus jagori*	Panay	ND2	A,B
AY922940	[8]	*Ptenochirus jagori*	Panay	ND2	A,B
AY922941	[8]	*Ptenochirus jagori*	Panay	ND2	A,B
AY922942	[8]	*Ptenochirus jagori*	Polillo	ND2	B,C
AY922943	[8]	*Ptenochirus jagori*	Polillo	ND2	B,C
AY922944	[8]	*Ptenochirus jagori*	Polillo	ND2	B,C
AY922945	[8]	*Ptenochirus jagori*	Polillo	ND2	B,C
AY922946	[8]	*Ptenochirus jagori*	Polillo	ND2	B,C
AY922947	[8]	*Ptenochirus jagori*	Polillo	ND2	B,C
AY922948	[8]	*Ptenochirus jagori*	Polillo	ND2	B,C
AY922949	[8]	*Ptenochirus jagori*	Polillo	ND2	B,C
AY922950	[8]	*Ptenochirus jagori*	Sibuyan	ND2	B
AY922951	[8]	*Ptenochirus jagori*	Sibuyan	ND2	B
AY922952	[8]	*Ptenochirus jagori*	Sibuyan	ND2	B
AY922953	[8]	*Ptenochirus jagori*	Siquijor	ND2	B,C
AY922954	[8]	*Ptenochirus jagori*	Siquijor	ND2	B,C
AY922955	[8]	*Ptenochirus jagori*	Siquijor	ND2	B,C
AY922956	[8]	*Ptenochirus jagori*	Siquijor	ND2	B,C
AY922957	[8]	*Ptenochirus jagori*	Siquijor	ND2	B,C
AY922958	[8]	*Ptenochirus jagori*	Siquijor	ND2	B,C
AY922959	[8]	*Ptenochirus jagori*	Siquijor	ND2	B,C
AY922960	[8]	*Ptenochirus jagori*	Siquijor	ND2	B,C
AY974611	[8]	*Ptenochirus minor*	Biliran	ND2	A,B
AY974612	[8]	*Ptenochirus minor*	Biliran	ND2	A,B
AY974613	[8]	*Ptenochirus minor*	Biliran	ND2	A,B
AY974614	[8]	*Ptenochirus minor*	Biliran	ND2	A,B
AY974615	[8]	*Ptenochirus minor*	Biliran	ND2	A,B
AY974616	[8]	*Ptenochirus minor*	Biliran	ND2	A,B
AY974617	[8]	*Ptenochirus minor*	Biliran	ND2	A,B
AY974618	[8]	*Ptenochirus minor*	Biliran	ND2	A,B
AY974619	[8]	*Ptenochirus minor*	Biliran	ND2	A,B
AY974627	[8]	*Ptenochirus minor*	Leyte	ND2	B
AY974628	[8]	*Ptenochirus minor*	Leyte	ND2	B
AY974629	[8]	*Ptenochirus minor*	Leyte	ND2	B
AY974630	[8]	*Ptenochirus minor*	Leyte	ND2	B
AY974631	[8]	*Ptenochirus minor*	Leyte	ND2	B
AY974632	[8]	*Ptenochirus minor*	Leyte	ND2	B
AY974633	[8]	*Ptenochirus minor*	Leyte	ND2	B
AY974620	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974621	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974622	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974623	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974624	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974625	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974626	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974634	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974635	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974636	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974637	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974638	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974639	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
AY974640	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974641	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974642	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974643	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974644	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974645	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974646	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974647	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974648	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974649	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974650	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974651	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974652	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974653	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974654	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974655	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
AY974656	[8]	*Ptenochirus minor*	Mindanao	ND2	A,B
JQ915565	KU 166455	*Hipposideros obscurus*	Leyte	ND2	A,B
JQ915566	KU 166456	*Hipposideros obscurus*	Leyte	ND2	A,B
JQ915567	KU 166457	*Hipposideros obscurus*	Leyte	ND2	A,B
JQ915568	KU 166458	*Hipposideros obscurus*	Leyte	ND2	A,B
JQ915569	KU 166459	*Hipposideros obscurus*	Leyte	ND2	A,B
JQ915570	KU 166460	*Hipposideros obscurus*	Leyte	ND2	A,B
JQ915571	KU 166461	*Hipposideros obscurus*	Leyte	ND2	A,B
JQ915572	KU 166462	*Hipposideros obscurus*	Leyte	ND2	A,B
JQ915573	KU 166463	*Hipposideros obscurus*	Leyte	ND2	A,B
JQ915507	FMNH 177464	*Hipposideros obscurus*	Luzon	ND2	C
JQ915508	FMNH 177465	*Hipposideros obscurus*	Luzon	ND2	C
JQ915509	FMNH 180196	*Hipposideros obscurus*	Luzon	ND2	C
JQ915510	FMNH 180197	*Hipposideros obscurus*	Luzon	ND2	C
JQ915511	FMNH 180198	*Hipposideros obscurus*	Luzon	ND2	C
JQ915512	FMNH 180199	*Hipposideros obscurus*	Luzon	ND2	C
JQ915513	FMNH 180200	*Hipposideros obscurus*	Luzon	ND2	C
JQ915514	FMNH 180201	*Hipposideros obscurus*	Luzon	ND2	C
JQ915515	FMNH 180285	*Hipposideros obscurus*	Luzon	ND2	C
JQ915516	FMNH 183313	*Hipposideros obscurus*	Luzon	ND2	C
JQ915517	FMNH 183314	*Hipposideros obscurus*	Luzon	ND2	C
JQ915518	FMNH 183315	*Hipposideros obscurus*	Luzon	ND2	C
JQ915519	FMNH 183316	*Hipposideros obscurus*	Luzon	ND2	C
JQ915520	FMNH 183317	*Hipposideros obscurus*	Luzon	ND2	C
JQ915521	FMNH 183318	*Hipposideros obscurus*	Luzon	ND2	C
JQ915522	FMNH 183320	*Hipposideros obscurus*	Luzon	ND2	C
JQ915523	FMNH 183321	*Hipposideros obscurus*	Luzon	ND2	C
JQ915524	FMNH 183322	*Hipposideros obscurus*	Luzon	ND2	C
JQ915525	FMNH 186034	*Hipposideros obscurus*	Luzon	ND2	C
JQ915526	FMNH 186035	*Hipposideros obscurus*	Luzon	ND2	C
JQ915527	FMNH 186036	*Hipposideros obscurus*	Luzon	ND2	C
JQ915529	FMNH 190050	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915530	FMNH 190051	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915531	FMNH 190052	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915532	FMNH 190053	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915533	FMNH 190054	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915534	FMNH 190055	*Hipposideros obscurus*	Mindanao	ND2	A,B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
JQ915535	FMNH 190057	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915536	FMNH 190058	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915537	FMNH 190059	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915538	FMNH 190060	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915539	FMNH 190061	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915540	FMNH 190062	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915541	FMNH 190063	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915542	FMNH 190064	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915543	FMNH 190065	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915544	FMNH 190066	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915545	FMNH 190067	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915546	FMNH 190068	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915547	FMNH 190069	*Hipposideros obscurus*	Mindanao	ND2	A,B
JQ915563	KU 166450	*Hipposideros obscurus*	Samar	ND2	B
JQ915564	KU 166454	*Hipposideros obscurus*	Samar	ND2	B
JQ915580	FMNH 202664	*Hipposideros pygmaeus*	Bohol	ND2	A,B
JQ915581	FMNH 202665	*Hipposideros pygmaeus*	Bohol	ND2	A,B
JQ915582	FMNH 202666	*Hipposideros pygmaeus*	Bohol	ND2	A,B
JQ915583	FMNH 202667	*Hipposideros pygmaeus*	Bohol	ND2	A,B
JQ915584	FMNH 202668	*Hipposideros pygmaeus*	Bohol	ND2	A,B
JQ915585	FMNH 202669	*Hipposideros pygmaeus*	Bohol	ND2	A,B
JQ915586	FMNH 202670	*Hipposideros pygmaeus*	Bohol	ND2	A,B
JQ915587	FMNH 202674	*Hipposideros pygmaeus*	Bohol	ND2	A,B
JQ915588	FMNH 202675	*Hipposideros pygmaeus*	Bohol	ND2	A,B
JQ915589	FMNH 202676	*Hipposideros pygmaeus*	Bohol	ND2	A,B
JQ915590	FMNH 202677	*Hipposideros pygmaeus*	Bohol	ND2	A,B
JQ915591	FMNH 202678	*Hipposideros pygmaeus*	Bohol	ND2	A,B
JQ915578	FMNH 190070	*Hipposideros pygmaeus*	Mindanao	ND2	A,B
JQ915579	FMNH 190071	*Hipposideros pygmaeus*	Mindanao	ND2	A,B
JQ915617	MSU 7[7]	*Hipposideros pygmaeus*	Mindanao	ND2	A,B
JQ915593	KU 164542	*Hipposideros pygmaeus*	Polillo	ND2	C
JQ915594	KU 164543	*Hipposideros pygmaeus*	Polillo	ND2	C
JQ915595	KU 164544	*Hipposideros pygmaeus*	Polillo	ND2	C
JQ915596	KU 164545	*Hipposideros pygmaeus*	Polillo	ND2	C
JQ915597	KU 164546	*Hipposideros pygmaeus*	Polillo	ND2	C
FJ813981/FJ814546	KU 165704	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813986/FJ814551	KU 165710	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813991/FJ814556	KU 165715	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813992/FJ814557	KU 165716	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813993/FJ814558	KU 165717	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813980/FJ814545	KU 165703	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813982/FJ814547	KU 165705	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813983/FJ814548	KU 165706	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813984/FJ814549	KU 165707	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813985/FJ814550	KU 165708	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813987/FJ814552	KU 165711	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813988/FJ814553	KU 165712	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813989/FJ814554	KU 165713	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813990/FJ814555	KU 165714	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813994/FJ814559	KU 165718	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ813995/FJ814560	KU 165720	*Crocidura beatus*	Camiguin	CYTB/ND2	C
FJ814007/FJ814572	KU 165756	*Crocidura beatus*	Leyte	CYTB/ND2	A,B
FJ814008/FJ814573	KU 165757	*Crocidura beatus*	Leyte	CYTB/ND2	A,B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
FJ814009/FJ814574	KU 165758	*Crocidura beatus*	Leyte	CYTB/ND2	A,B
FJ814010/FJ814575	KU 165759	*Crocidura beatus*	Leyte	CYTB/ND2	A,B
FJ814011/FJ814576	KU 165760	*Crocidura beatus*	Leyte	CYTB/ND2	A,B
FJ814012/FJ814577	KU 165761	*Crocidura beatus*	Leyte	CYTB/ND2	A,B
FJ814013/FJ814578	KU 165762	*Crocidura beatus*	Leyte	CYTB/ND2	A,B
FJ814014/FJ814579	KU 165763	*Crocidura beatus*	Leyte	CYTB/ND2	A,B
FJ814015/FJ814580	KU 165766	*Crocidura beatus*	Leyte	CYTB/ND2	A,B
FJ814016/FJ814581	KU 165767	*Crocidura beatus*	Leyte	CYTB/ND2	A,B
FJ814017/FJ814582	KU 165768	*Crocidura beatus*	Leyte	CYTB/ND2	A,B
FJ814018/FJ814583	KU 165775	*Crocidura beatus*	Leyte	CYTB/ND2	A,B
FJ813844/FJ814410	FMNH 146965	*Crocidura beatus*	Mindanao	CYTB/ND2	B
FJ813845/FJ814411	FMNH 146966	*Crocidura beatus*	Mindanao	CYTB/ND2	B
FJ813846/FJ814412	FMNH 147819	*Crocidura beatus*	Mindanao	CYTB/ND2	B
FJ813847/FJ814413	FMNH 166459	*Crocidura beatus*	Mindanao	CYTB/ND2	B
FJ813884/FJ814449	FMNH 191345	*Crocidura beatus*	Mindanao	CYTB/ND2	B
FJ814019/FJ814584	No voucher	*Crocidura beatus*	Mindanao	CYTB/ND2	B
FJ813996/FJ814561	KU 165742	*Crocidura beatus*	Samar	CYTB/ND2	A,B
FJ813997/FJ814562	KU 165744	*Crocidura beatus*	Samar	CYTB/ND2	A,B
FJ813998/FJ814563	KU 165745	*Crocidura beatus*	Samar	CYTB/ND2	A,B
FJ813999/FJ814564	KU 165746	*Crocidura beatus*	Samar	CYTB/ND2	A,B
FJ814000/FJ814565	KU 165748	*Crocidura beatus*	Samar	CYTB/ND2	A,B
FJ814001/FJ814566	KU 165749	*Crocidura beatus*	Samar	CYTB/ND2	A,B
FJ814002/FJ814567	KU 165750	*Crocidura beatus*	Samar	CYTB/ND2	A,B
FJ814003/FJ814568	KU 165751	*Crocidura beatus*	Samar	CYTB/ND2	A,B
FJ814004/FJ814569	KU 165752	*Crocidura beatus*	Samar	CYTB/ND2	A,B
FJ814005/FJ814570	KU 165753	*Crocidura beatus*	Samar	CYTB/ND2	A,B
FJ814006/FJ814571	KU 165754	*Crocidura beatus*	Samar	CYTB/ND2	A,B
FJ813951/FJ814516	KU 165046	*Crocidura negrina*	Negros	CYTB/ND2	A,B
FJ813952/FJ814517	KU 165047	*Crocidura negrina*	Negros	CYTB/ND2	A,B
FJ813953/FJ814518	KU 165048	*Crocidura negrina*	Negros	CYTB/ND2	A,B
FJ813954/FJ814519	KU 165049	*Crocidura negrina*	Negros	CYTB/ND2	A,B
FJ813955/FJ814520	KU 165101	*Crocidura negrina*	Negros	CYTB/ND2	A,B
FJ813956/FJ814521	KU 165102	*Crocidura negrina*	Negros	CYTB/ND2	A,B
FJ813957/FJ814522	KU 165103	*Crocidura negrina*	Negros	CYTB/ND2	A,B
FJ813958/FJ814523	KU 165104	*Crocidura negrina*	Negros	CYTB/ND2	A,B
FJ813959/FJ814524	KU 165105	*Crocidura negrina*	Negros	CYTB/ND2	A,B
FJ813960/FJ814525	KU 165106	*Crocidura negrina*	Negros	CYTB/ND2	A,B
FJ813961/FJ814526	KU 165107	*Crocidura negrina*	Negros	CYTB/ND2	A,B
FJ813962/FJ814527	KU 165108	*Crocidura negrina*	Negros	CYTB/ND2	A,B
FJ813902/FJ814467	FMNH 195214	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813903/FJ814468	FMNH 195215	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813907/FJ814472	FMNH 195219	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813918/FJ814483	FMNH 195991	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813904/FJ814469	FMNH 195216	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813905/FJ814470	FMNH 195217	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813906/FJ814471	FMNH 195218	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813908/FJ814473	FMNH 195220	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813909/FJ814474	FMNH 195221	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813910/FJ814475	FMNH 195223	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813911/FJ814476	FMNH 195233	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813912/FJ814477	FMNH 195224	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813913/FJ814478	FMNH 195227	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813914/FJ814479	FMNH 195228	*Crocidura palawanensis*	Palawan	CYTB/ND2	C

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.

Table5.8—cont.

Genbank	Voucher ID	Species	Island	Gene	Use
FJ813915/FJ814480	FMNH 195229	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813916/FJ814481	FMNH 195230	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813917/FJ814482	FMNH 195231	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813919/FJ814484	FMNH 195992	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813920/FJ814485	FMNH 195993	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813921/FJ814486	FMNH 195994	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813922/FJ814487	FMNH 195995	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813923/FJ814488	FMNH 195996	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813978/FJ814543	KU 165463	*Crocidura palawanensis*	Palawan	CYTB/ND2	C
FJ813944/FJ814508	KU 164874	*Crocidura panayensis*	Panay	CYTB/ND2	A,B
FJ813945/FJ814509	KU 164875	*Crocidura panayensis*	Panay	CYTB/ND2	A,B
FJ813946/FJ814510	KU 164876	*Crocidura panayensis*	Panay	CYTB/ND2	A,B
FJ813947/FJ814511	KU 164877	*Crocidura panayensis*	Panay	CYTB/ND2	A,B
FJ813948/FJ814512	KU 164878	*Crocidura panayensis*	Panay	CYTB/ND2	A,B
FJ813950/FJ814515	KU 164993	*Crocidura panayensis*	Panay	CYTB/ND2	A,B

Uses: A = used in msBayes analyses; B = used in gene tree estimation; C = used for θ estimates.
Collection abbreviations: CMNH = Cincinnati Museum of Natural History; FMNH = Field Museum of Natural History; KU = University of Kansas Natural History Museum; PNM = National Museum of the Philippines; TNHC = Texas Natural History Collection of the University of Texas at Austin; USNM = National Museum of Natural History.
[1] Rafe M. Brown field catalog, deposited in PNM.
[2] Arvin C. Diesmos field catalog, deposited in PNM.
[3] Genevieve V. A. Gee field catalog, deposited in PNM.
[4] Elsa Dolima field catalog, deposited in PNM.
[5] PNM/CMNH Philippine Biodiversity Inventory Herpetology field collection (deposited in PNM).
[6] Cameron D. Siler field series, non-vouchered genetic sample deposited at KU.
[7] Non-vouchered genetic sample from Mindanao State University.
[8] Downloaded from Genbank; Please see Genbank for voucher information.

5.12 Figures

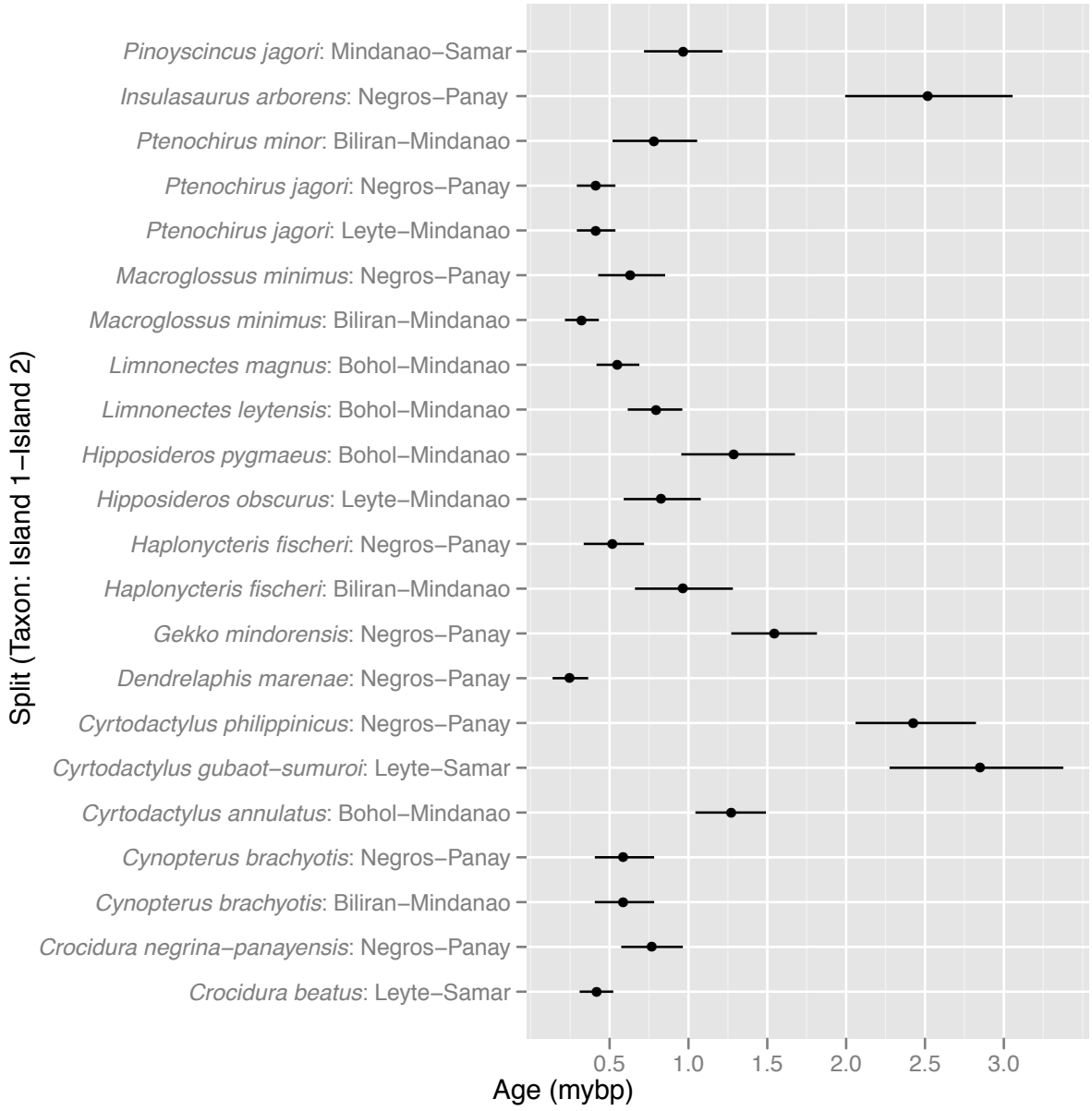

Figure S5.2. Plot of the mean and 95% highest posterior density of gene divergence times estimated for each of the 22 population pairs using a fixed rate of 2×10^{-8} substitutions per site per year in BEAST.

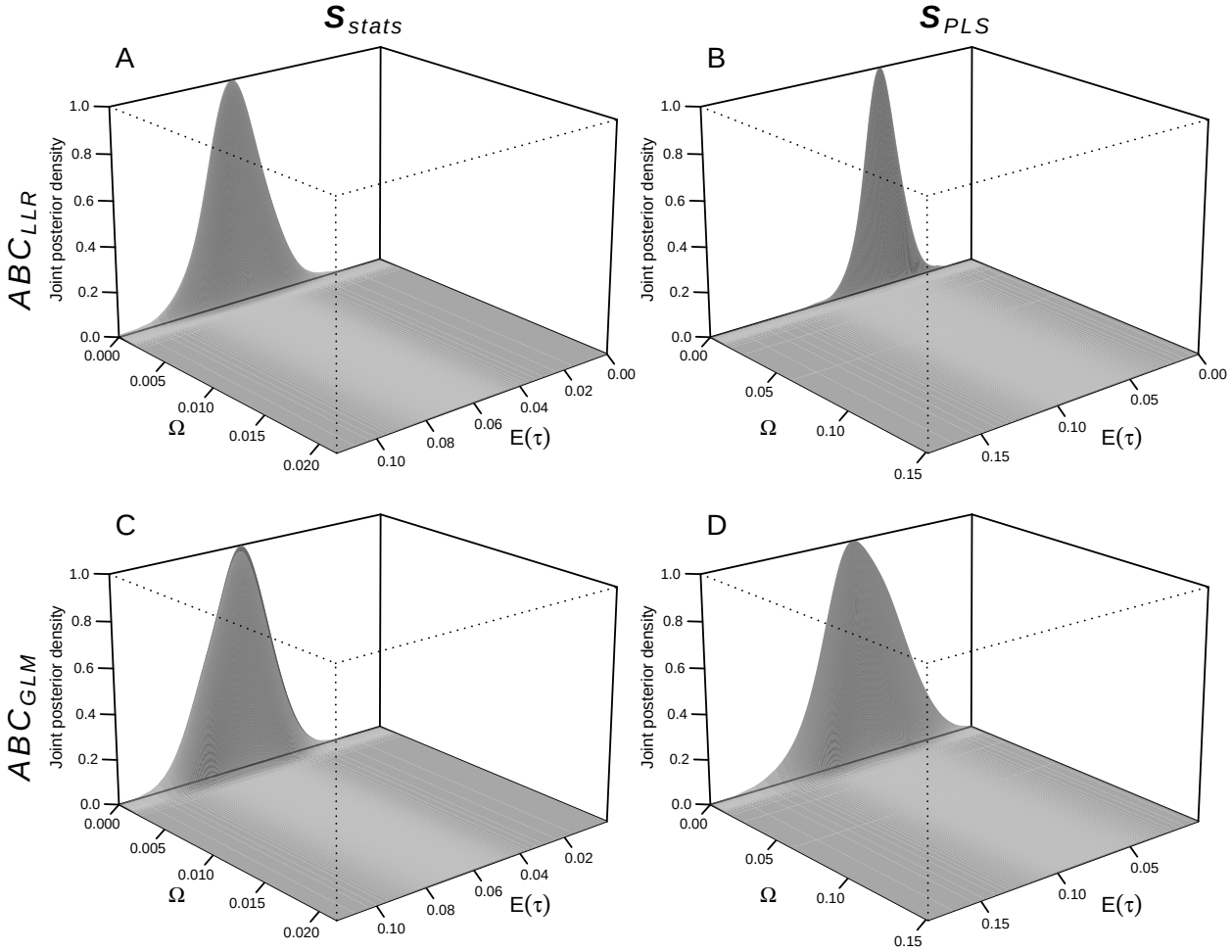

Figure S5.3. The estimated joint posterior densities of the dispersion index (Ω) and mean ($E(\tau)$) of divergence time vector, $\boldsymbol{\tau}$, using (A & B) ABC_{LLR} and (C & D) ABC_{GLM} regression methods, and (A & C) the `msBayes` summary statistics and (B & D) partial least squares (PLS) components of the `msBayes` summary statistics. Ω and $E(\tau)$ are in units of N_C generations. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

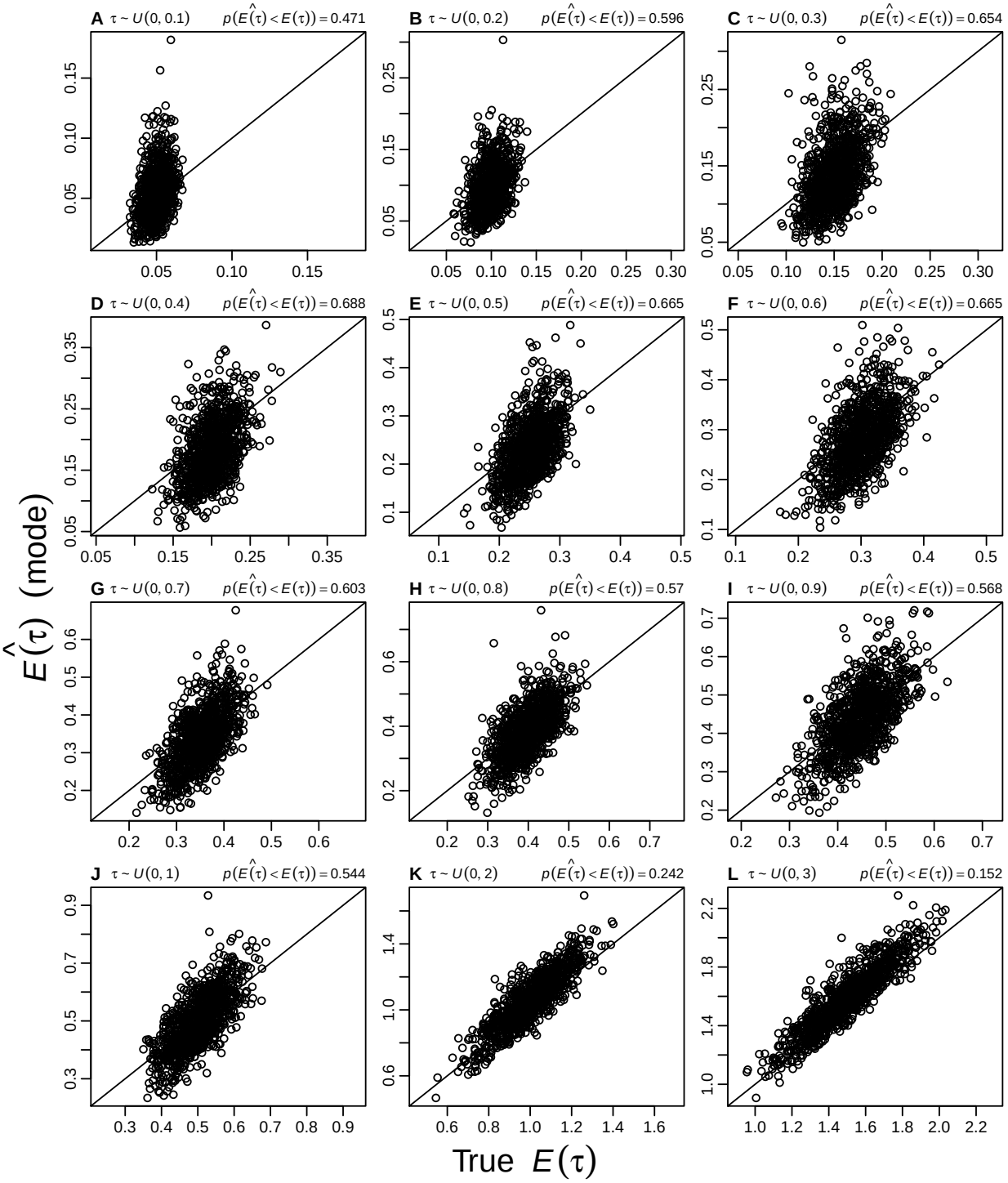

Figure S5.4. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

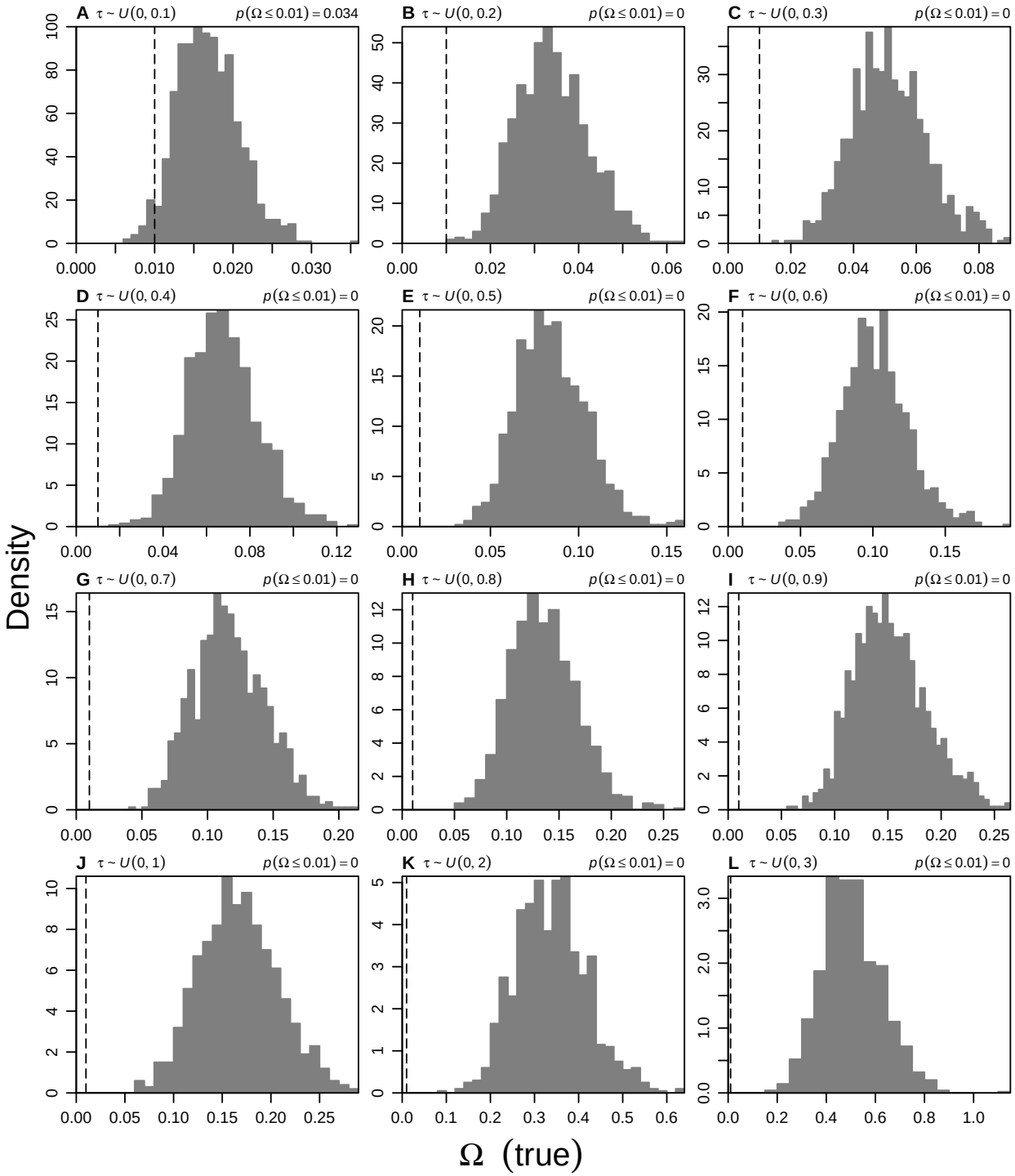

Figure S5.5. Histograms of the true dispersion index of divergence times (Ω) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the proportion of true Ω values consistent with one divergence event, $p(\Omega \leq 0.01)$, is given for each τ_{max}.

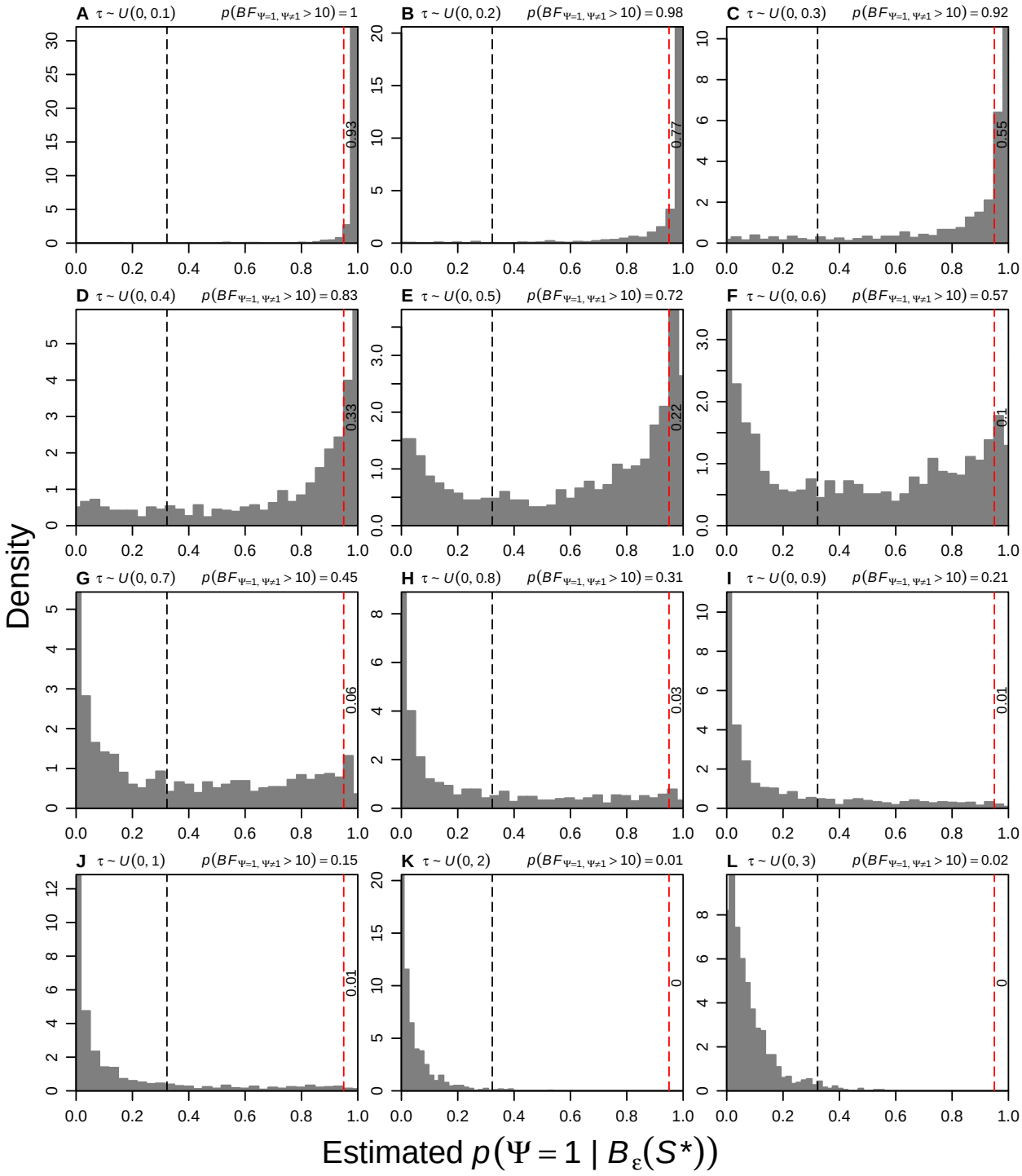

Figure S5.6. Histograms of the estimated posterior probability of one divergence event, $p(\Psi = 1|B_\epsilon(\mathbf{S}^*))$, from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event with a Bayes factor greater than 10 (dashed black line), $p(BF_{\Psi=1,\Psi\neq1} > 10)$, is given for each τ_{max}. The red line indicates $p(\Psi = 1|B_\epsilon(\mathbf{S}^*)) = 0.95$, and the estimated probability of inferring a posterior probability greater than 0.95 is given to the right of the line.. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

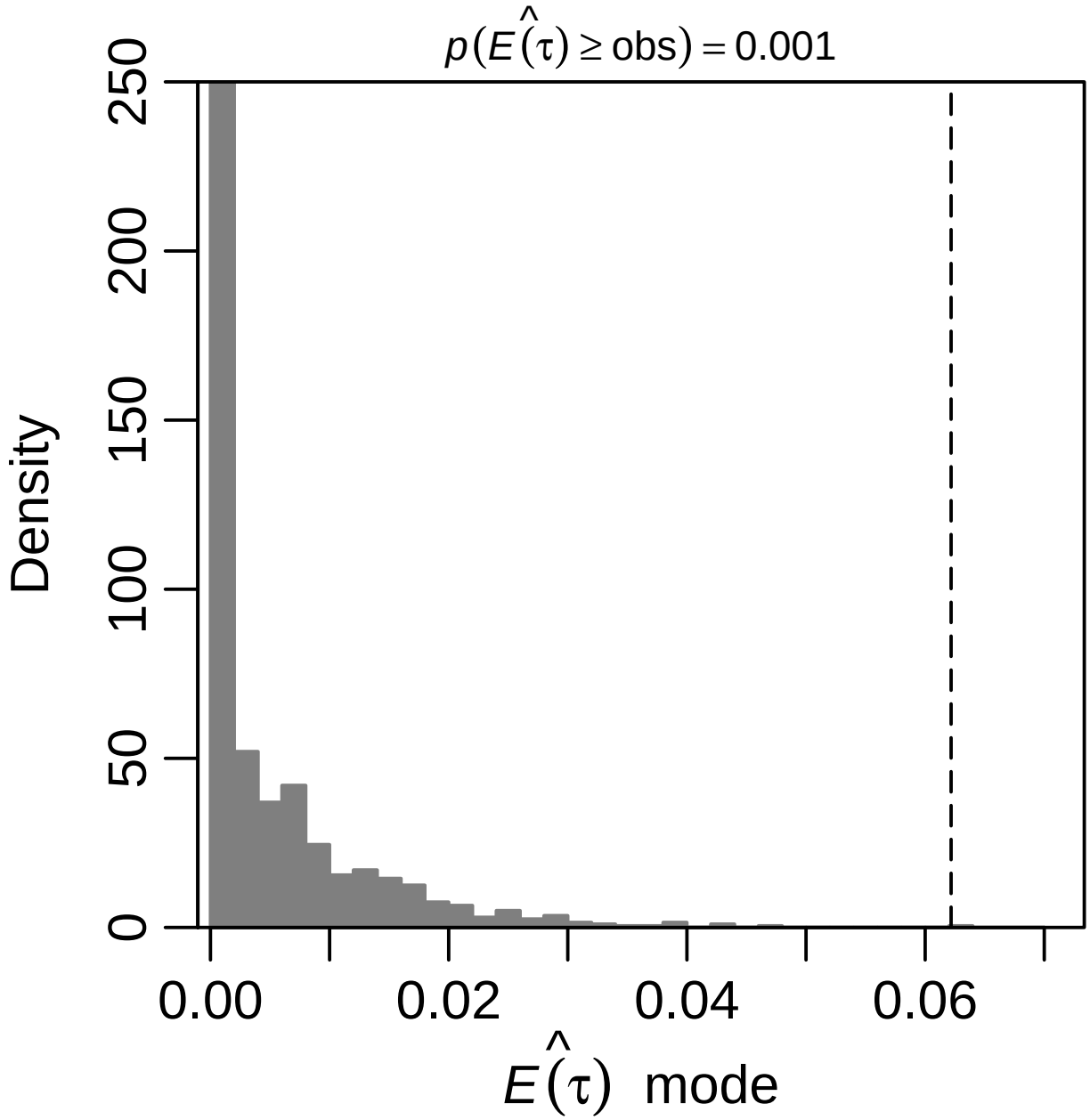

Figure S5.7. Results of 1000 simulation replicates of no divergence between population pairs, using $\mathbb{S}_{stats}$, ABC_{GLM}, and 5×10^6 samples from the prior. The posterior mode estimate of the mean of the divergence time vector, $E(\tau)$ (in $4N_C$ generations), is plotted for each replicate, and the observed estimate from the empirical data is indicated by the dotted line. The proportion of the simulation replicates that exceeded the observed empirical estimate is also given.

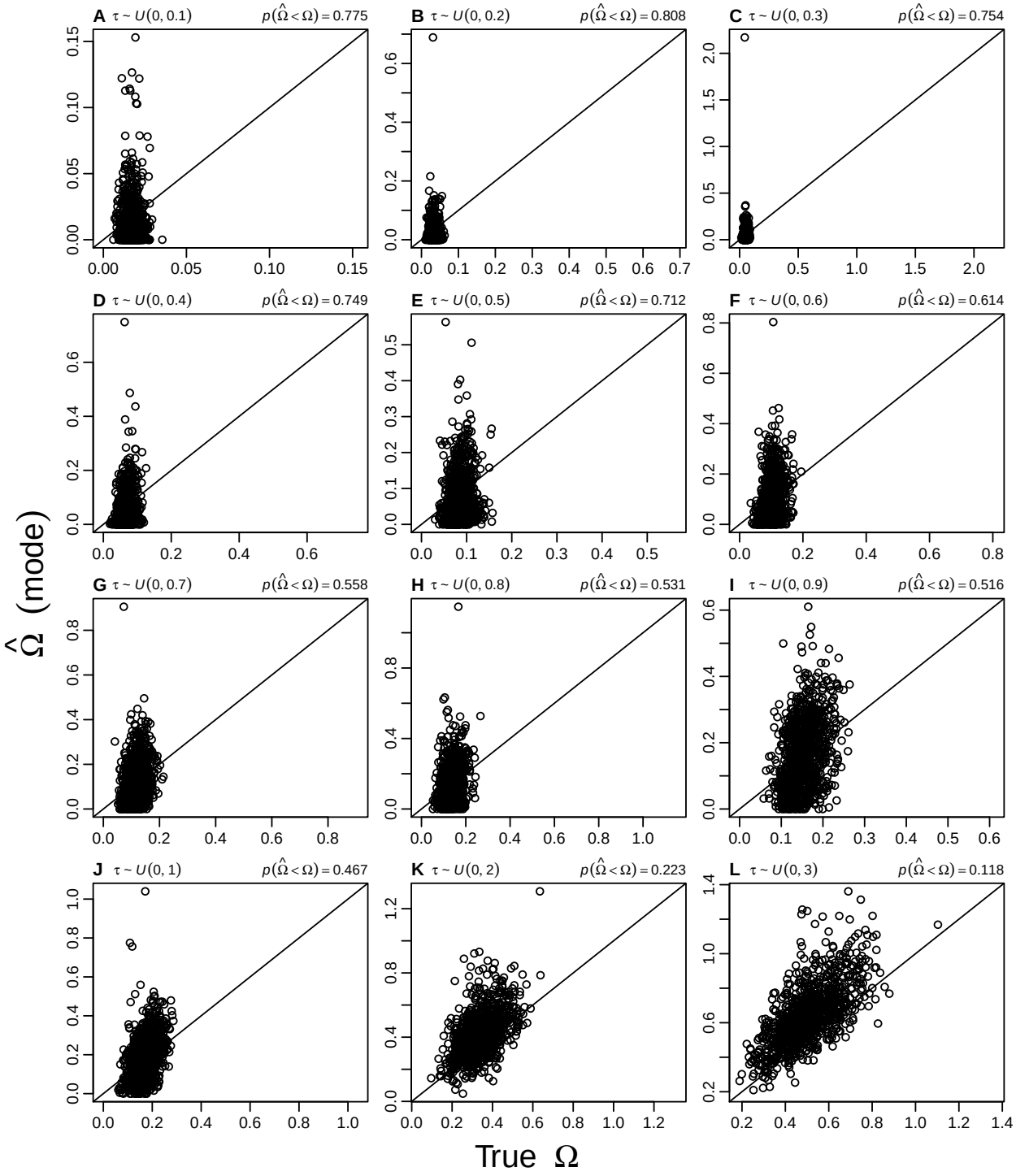

Figure S5.8. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

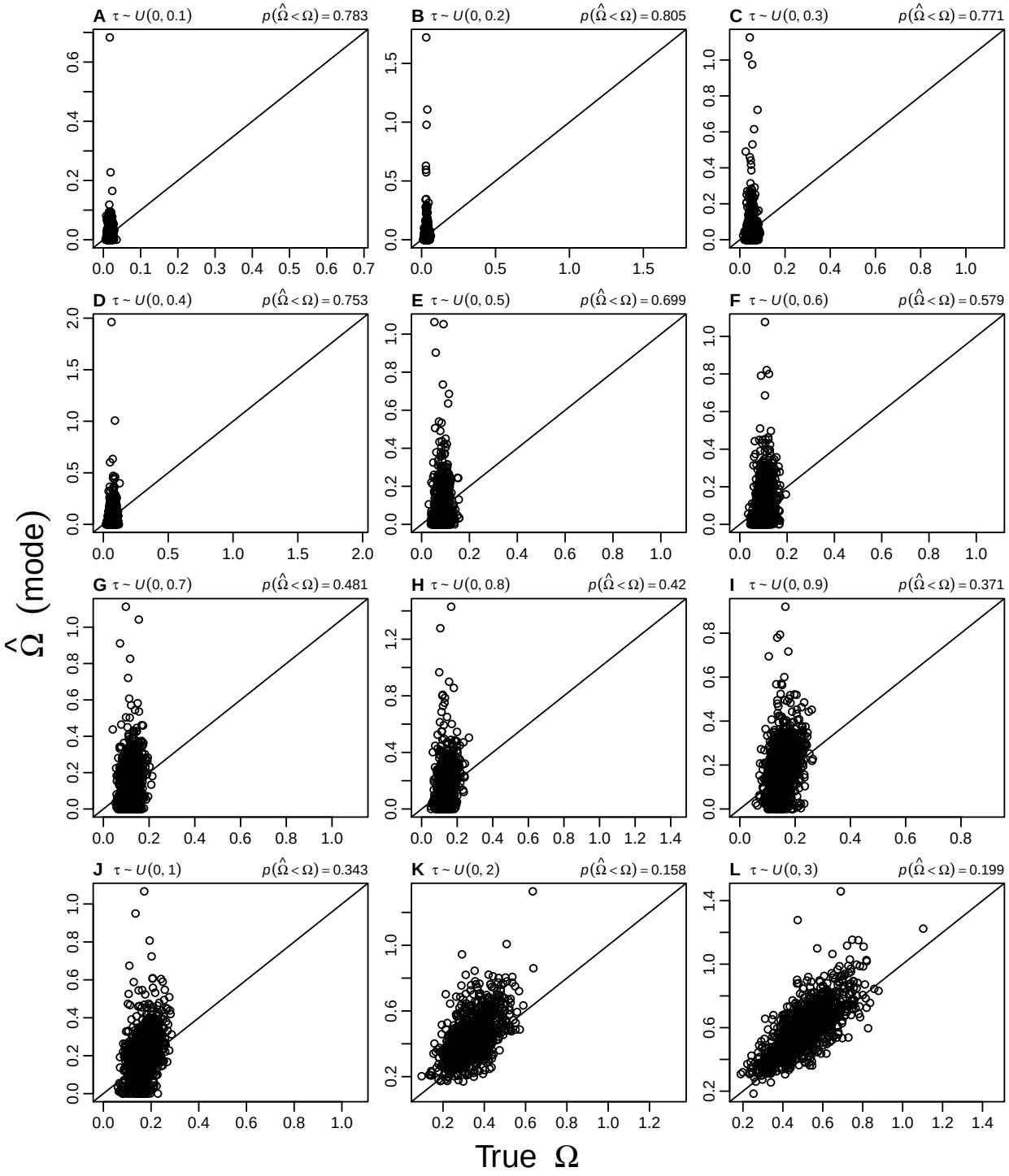

Figure S5.9. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

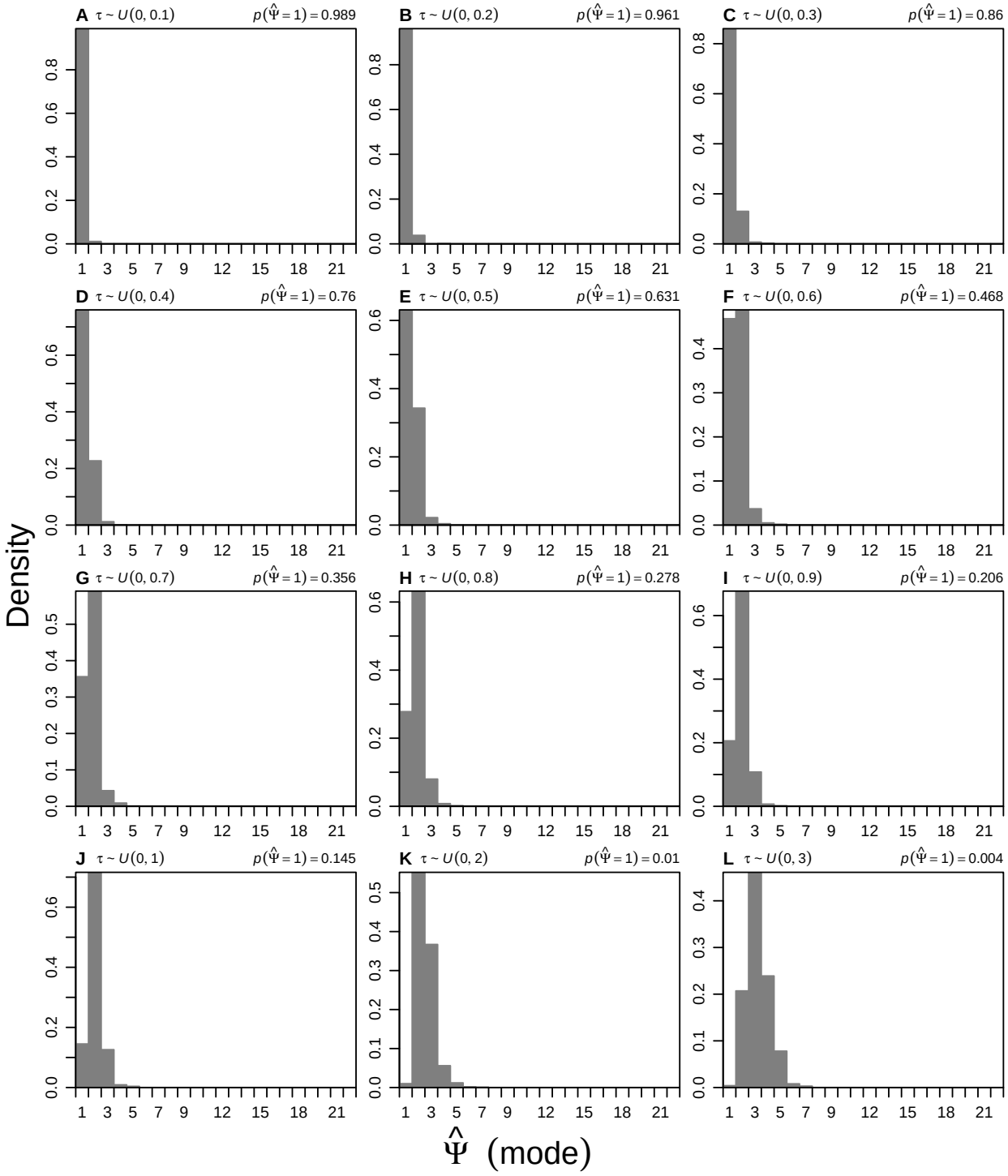

Figure S5.10. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

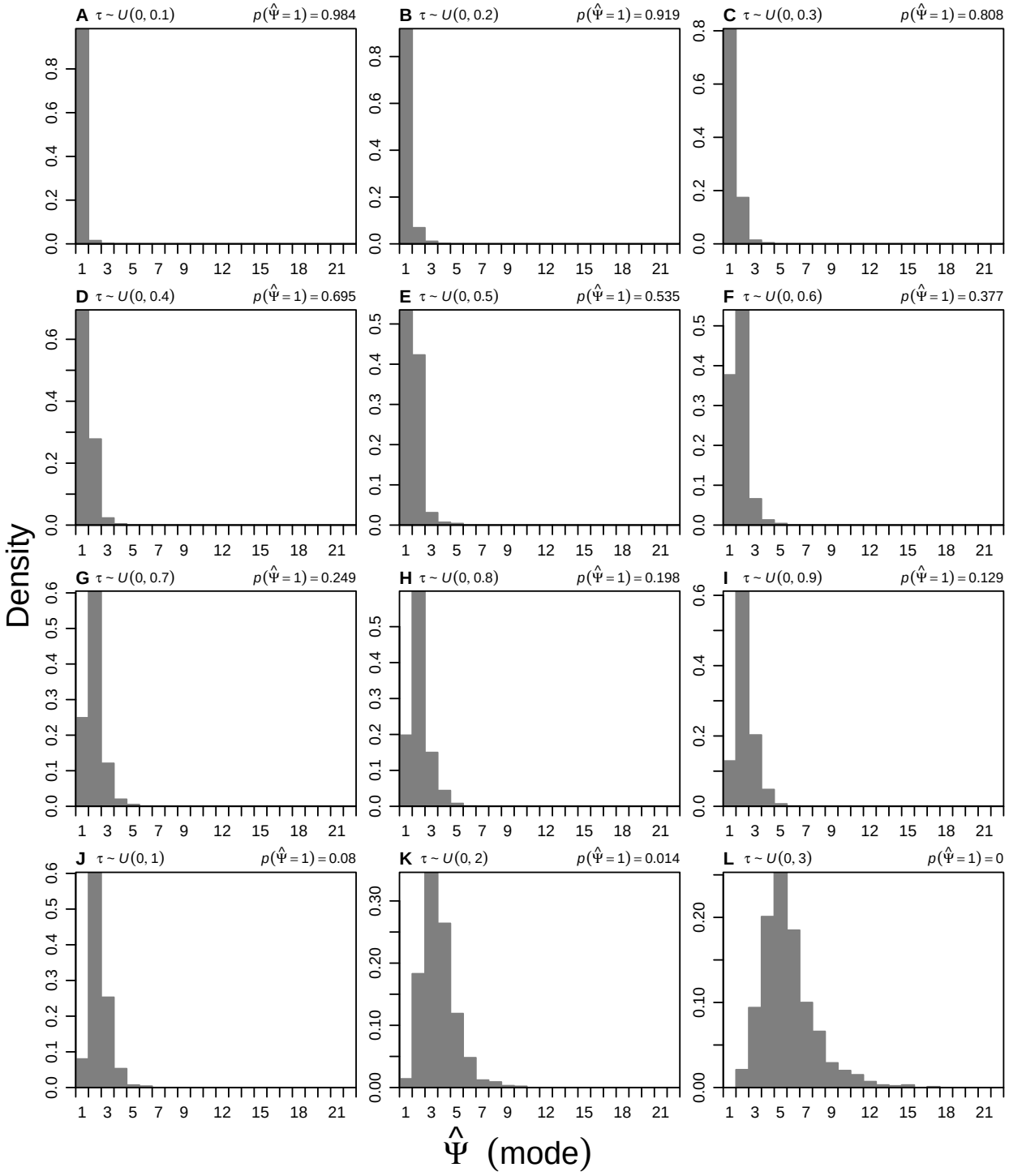

Figure S5.11. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

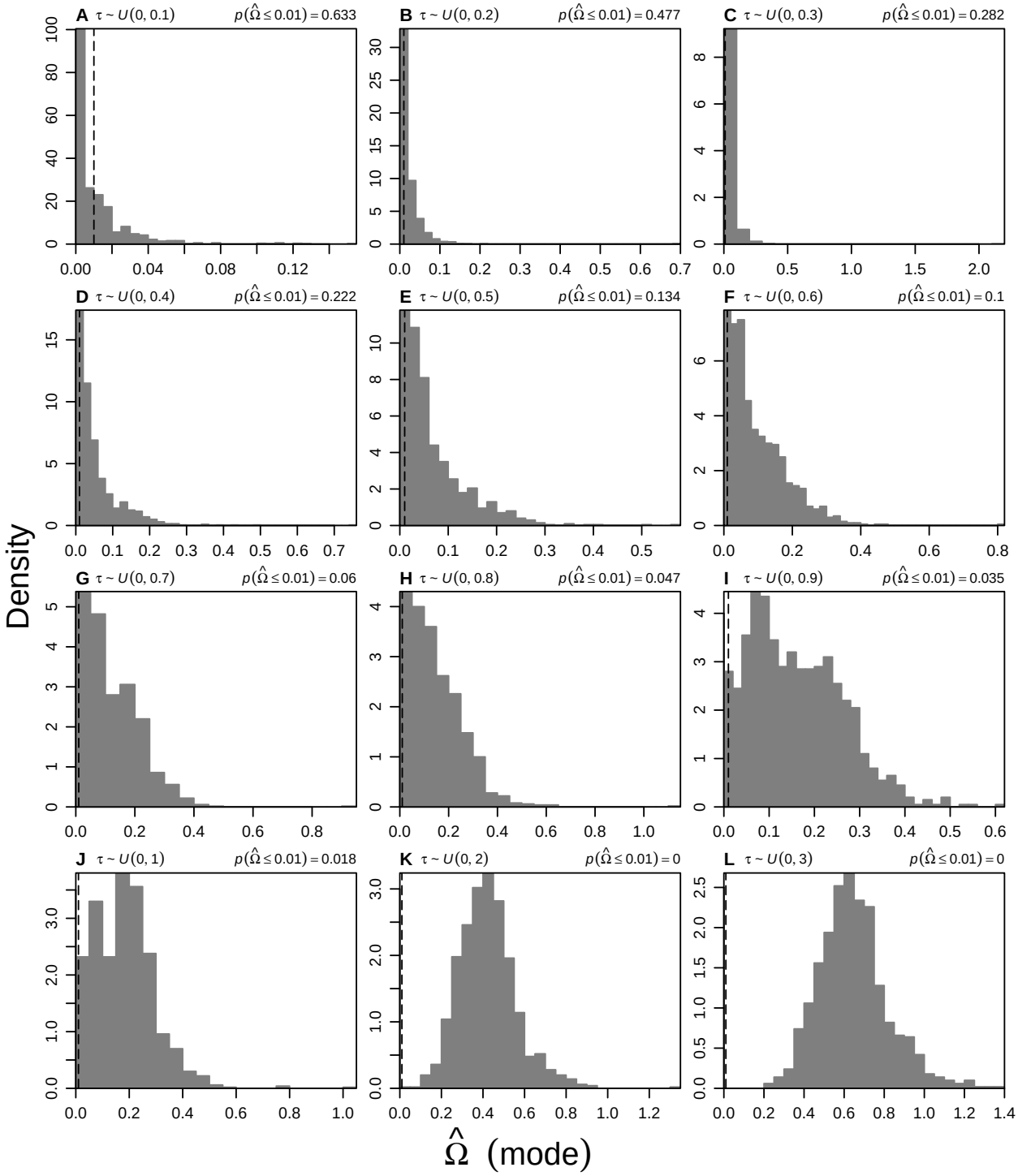

Figure S5.12. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

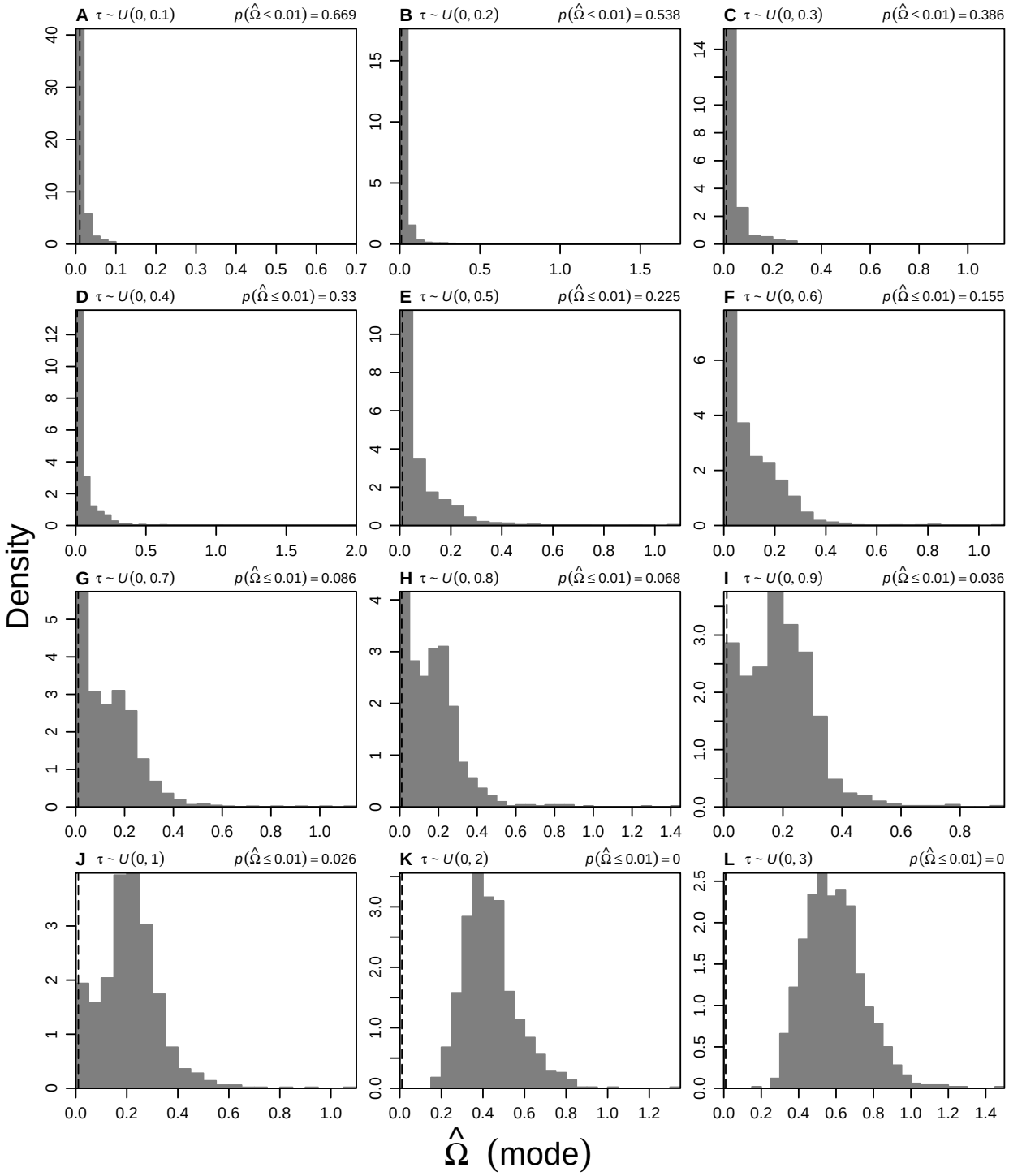

Figure S5.13. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

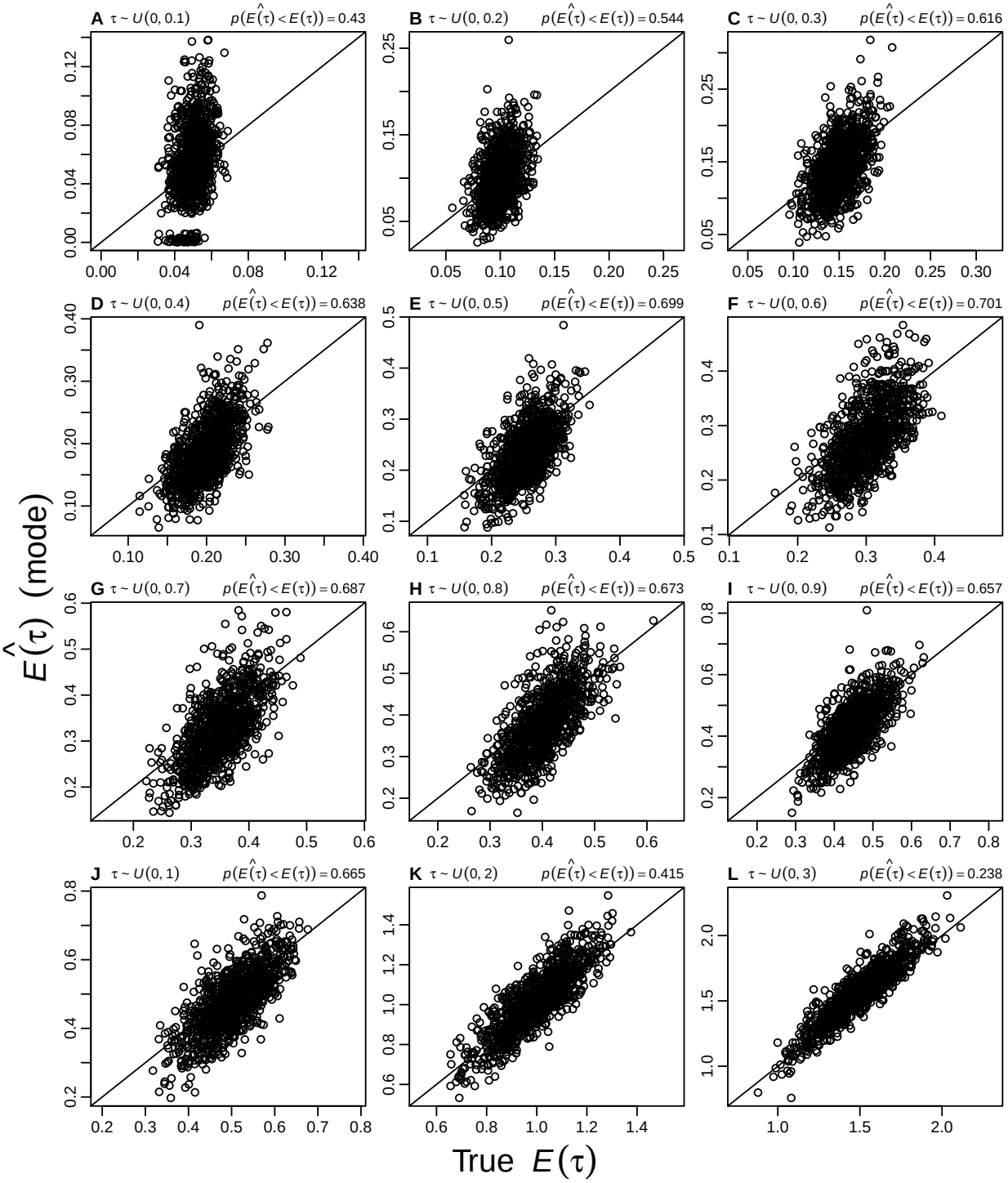

Figure S5.14. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

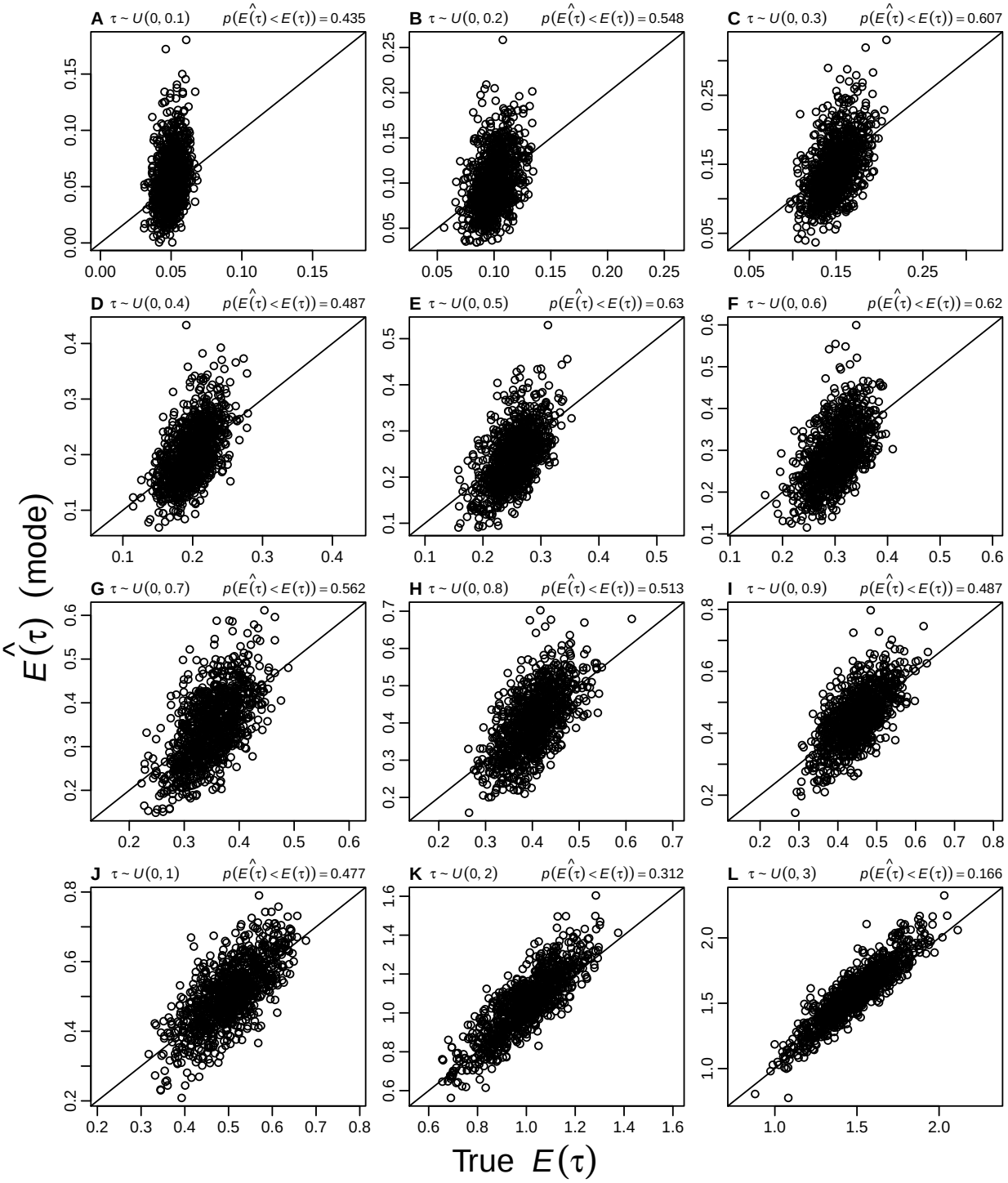

Figure S5.15. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

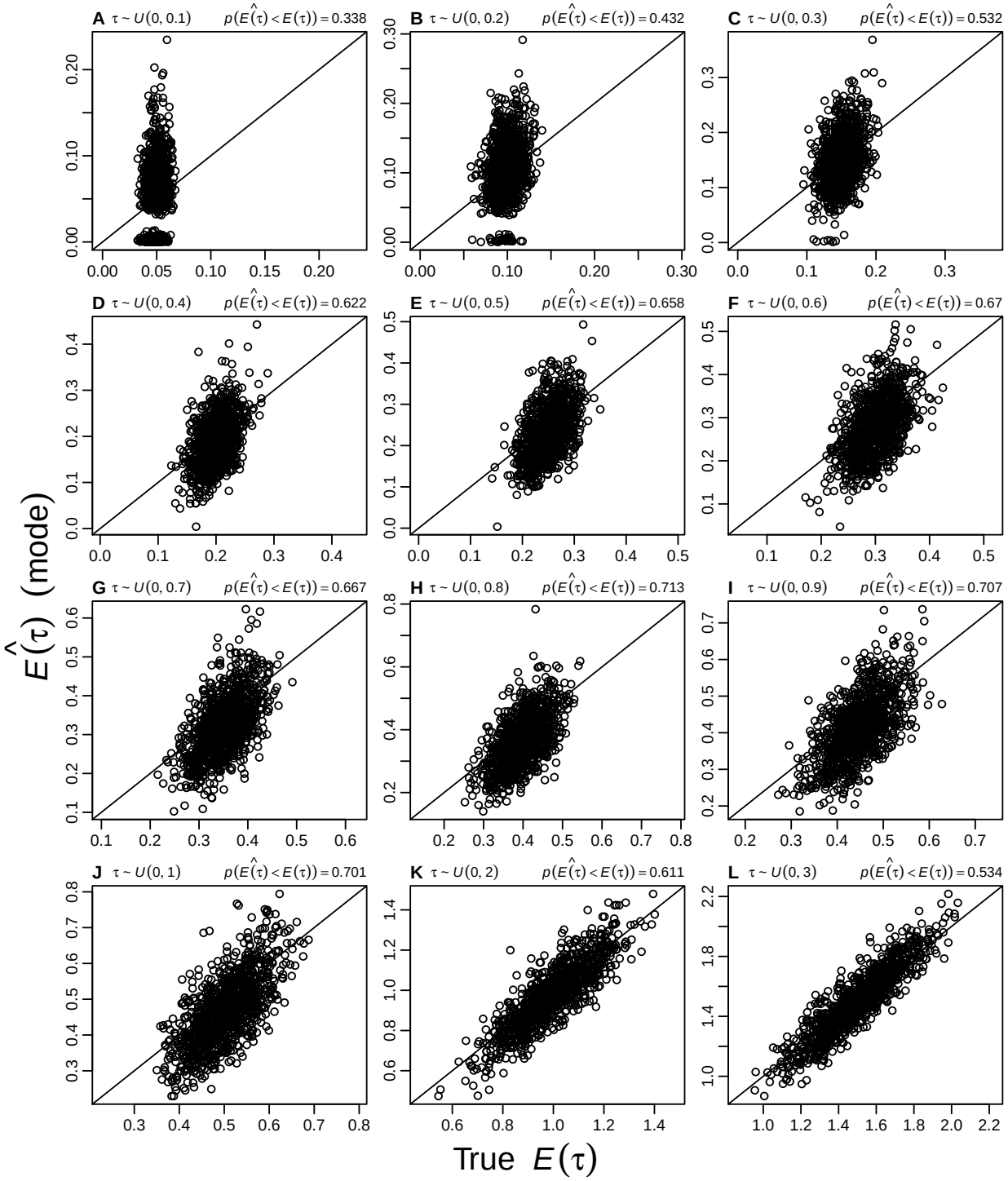

Figure S5.16. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

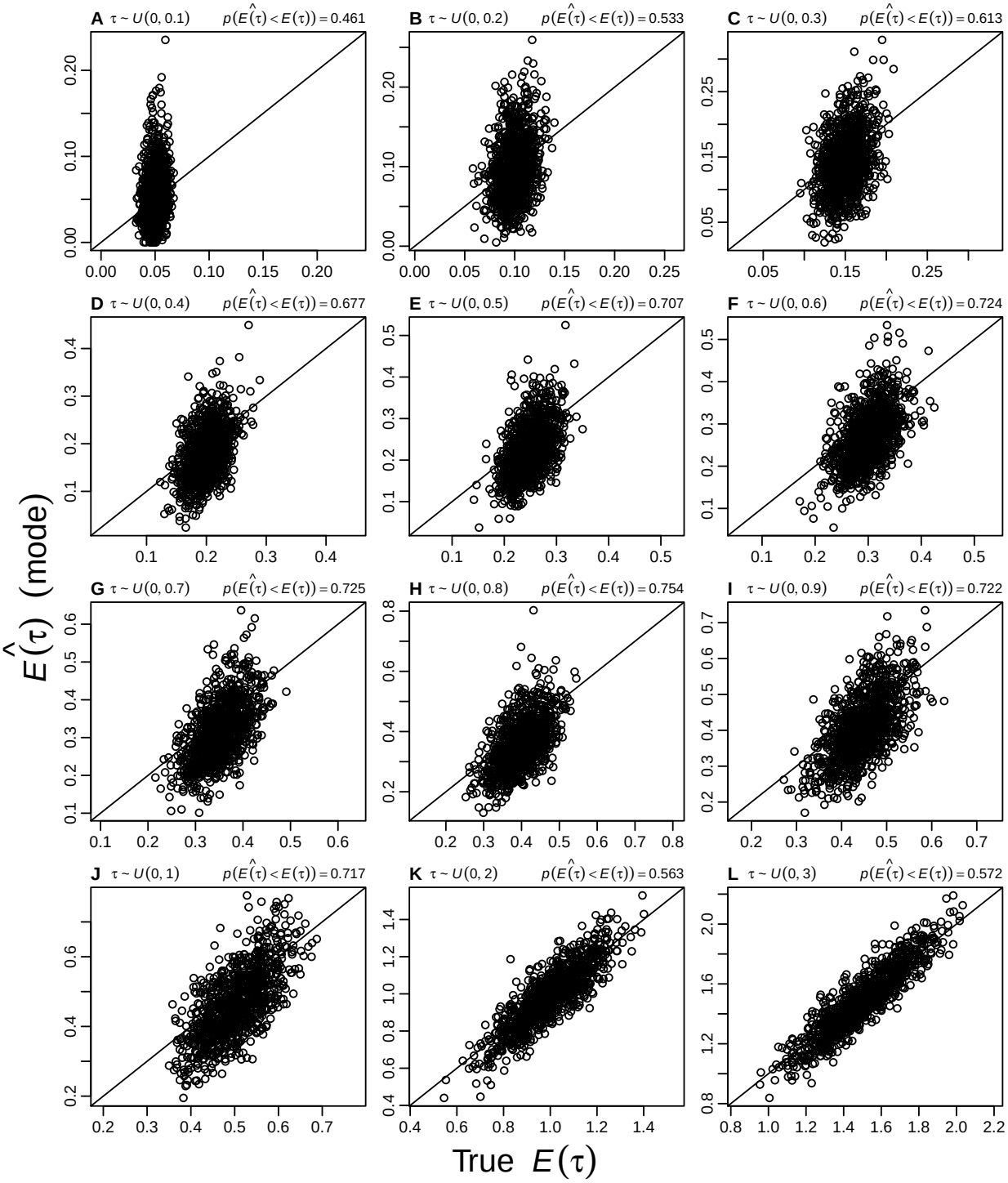

Figure S5.17. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

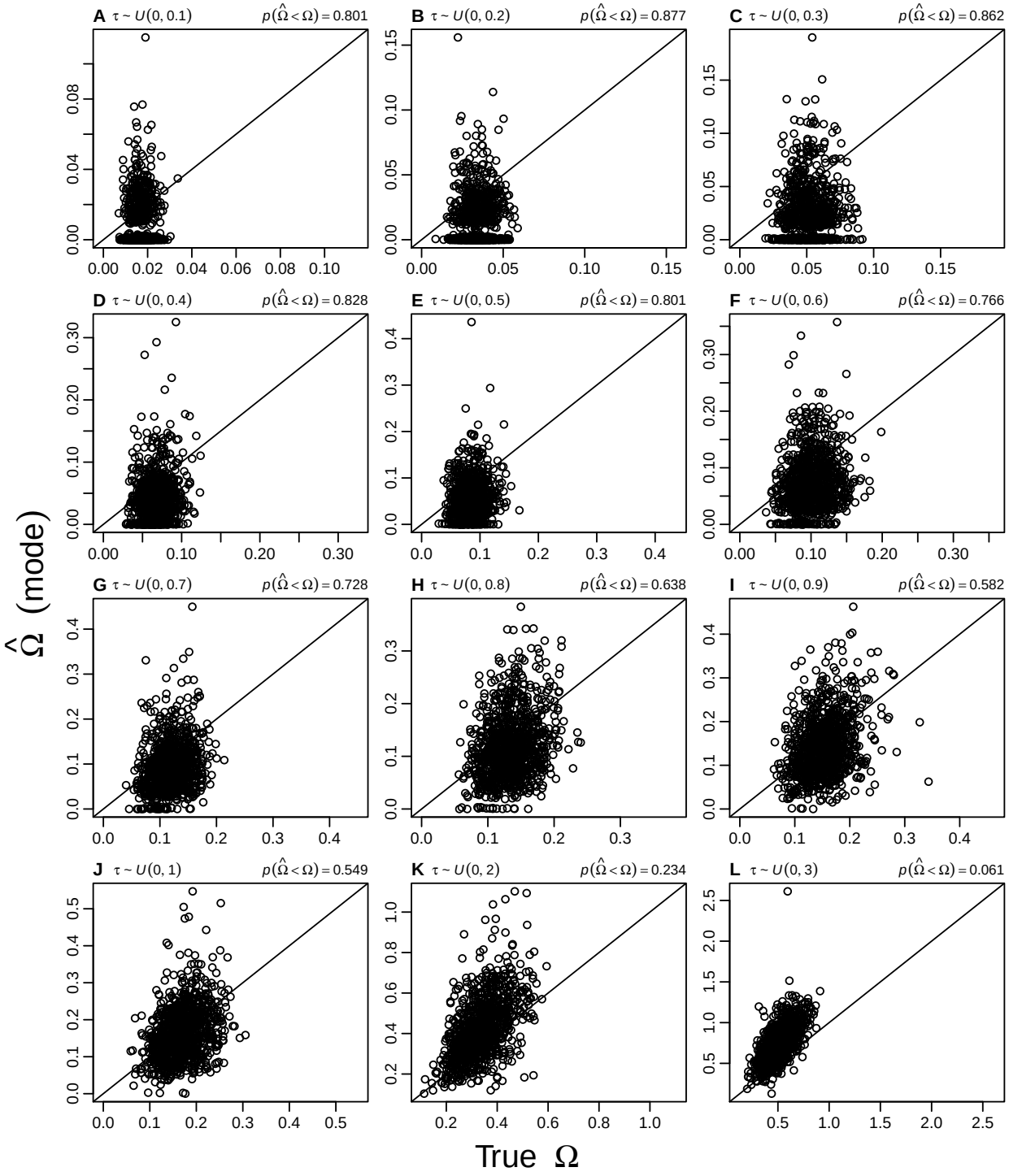

Figure S5.18. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

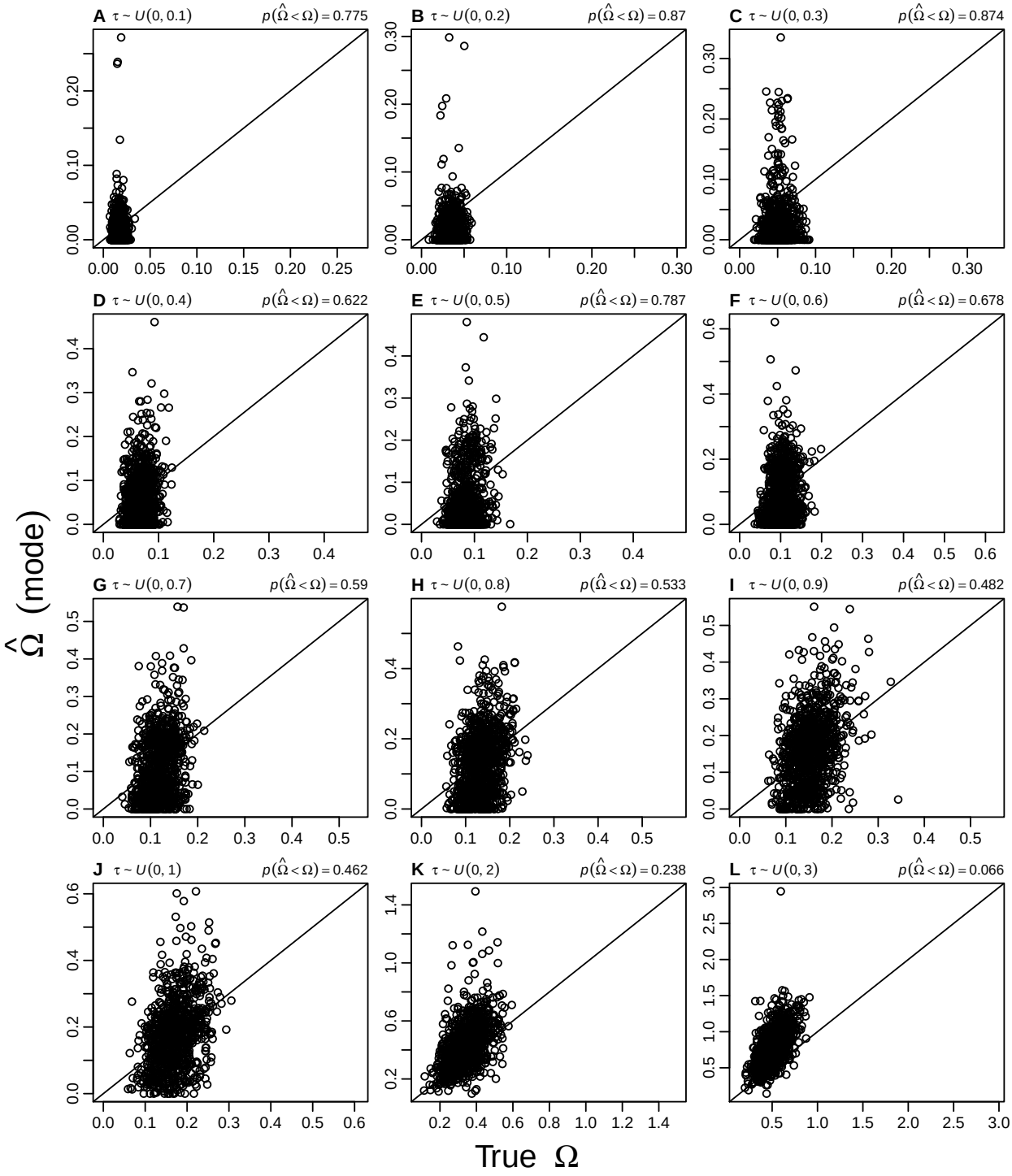

Figure S5.19. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

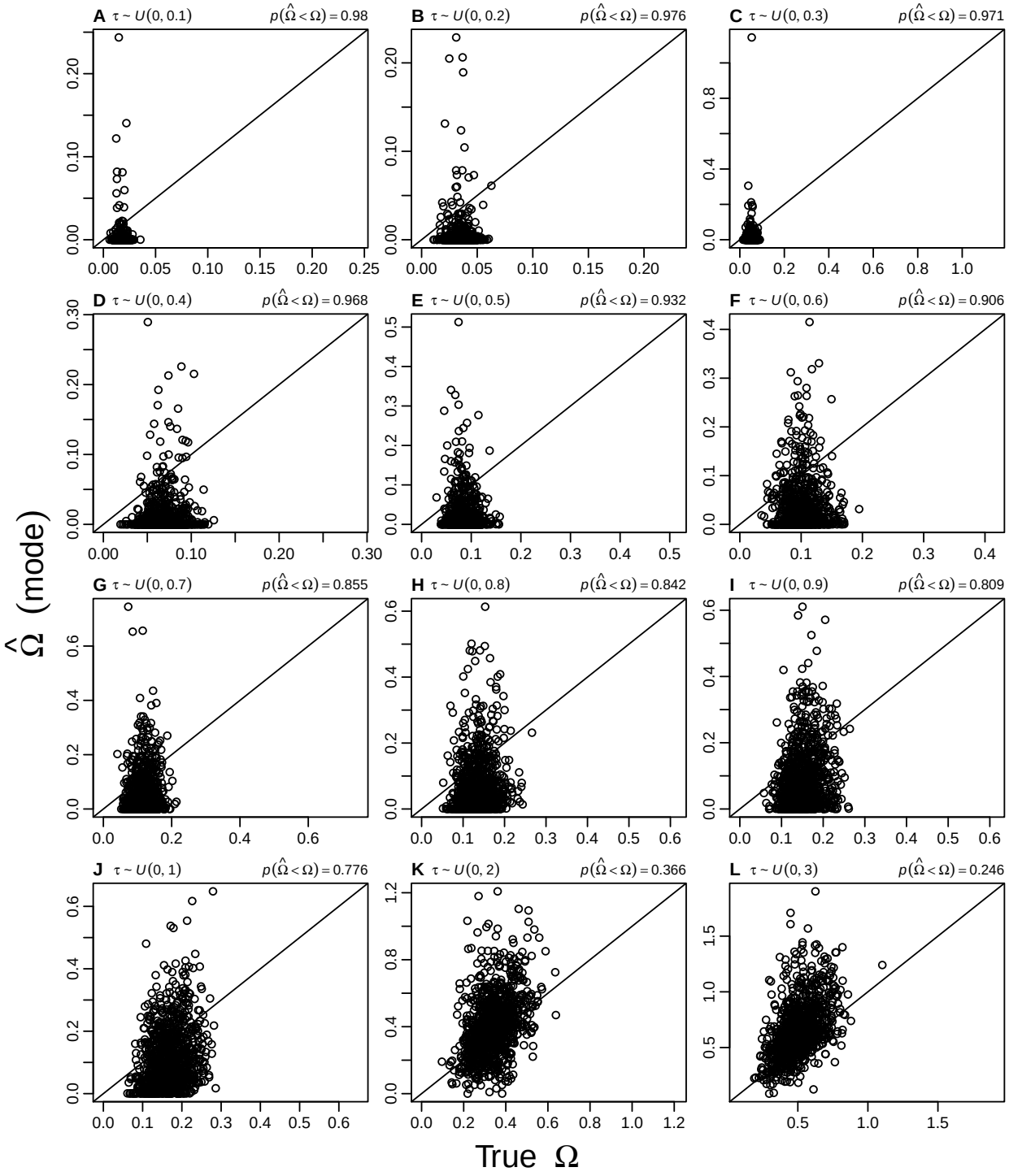

Figure S5.20. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

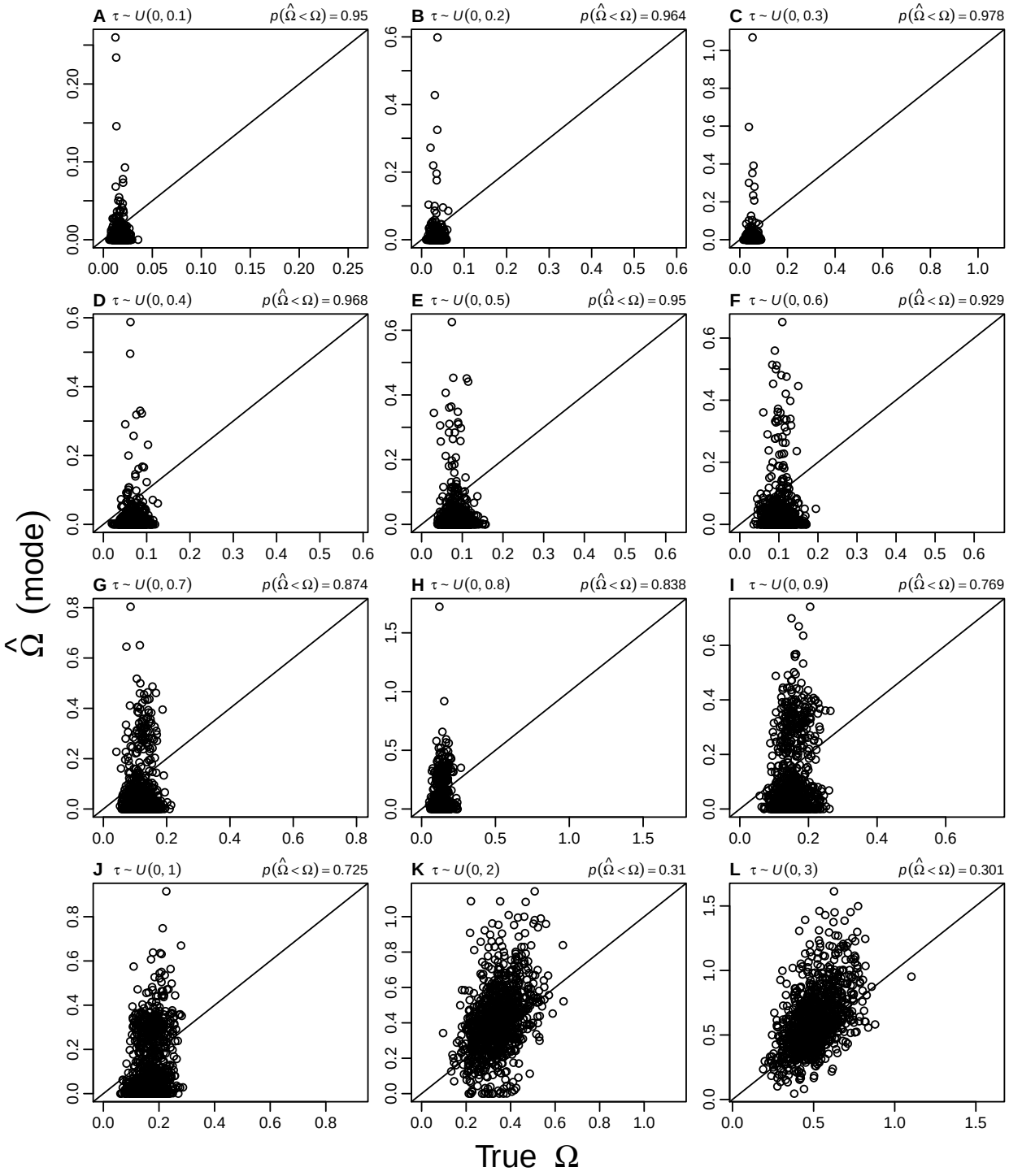

Figure S5.21. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

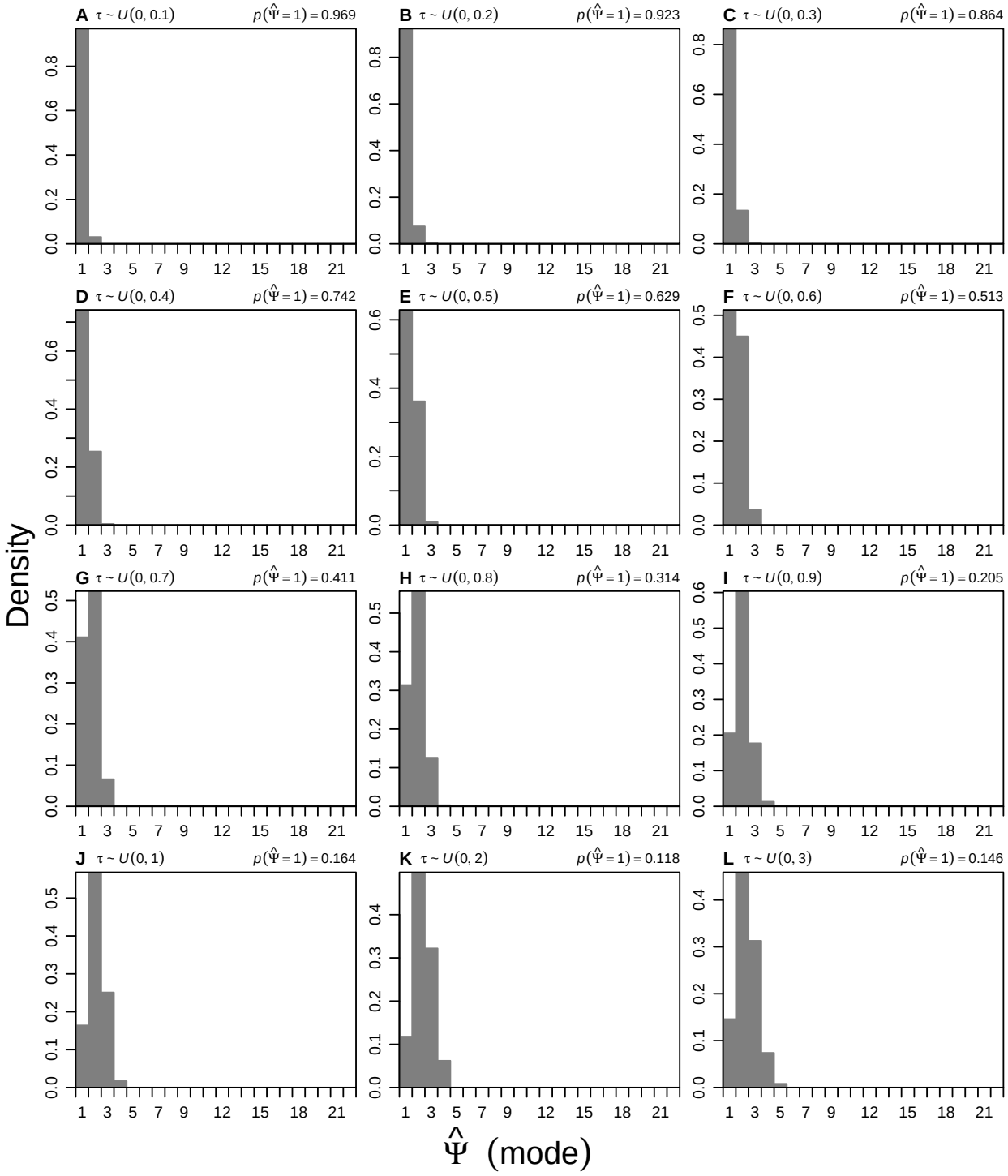

Figure S5.22. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

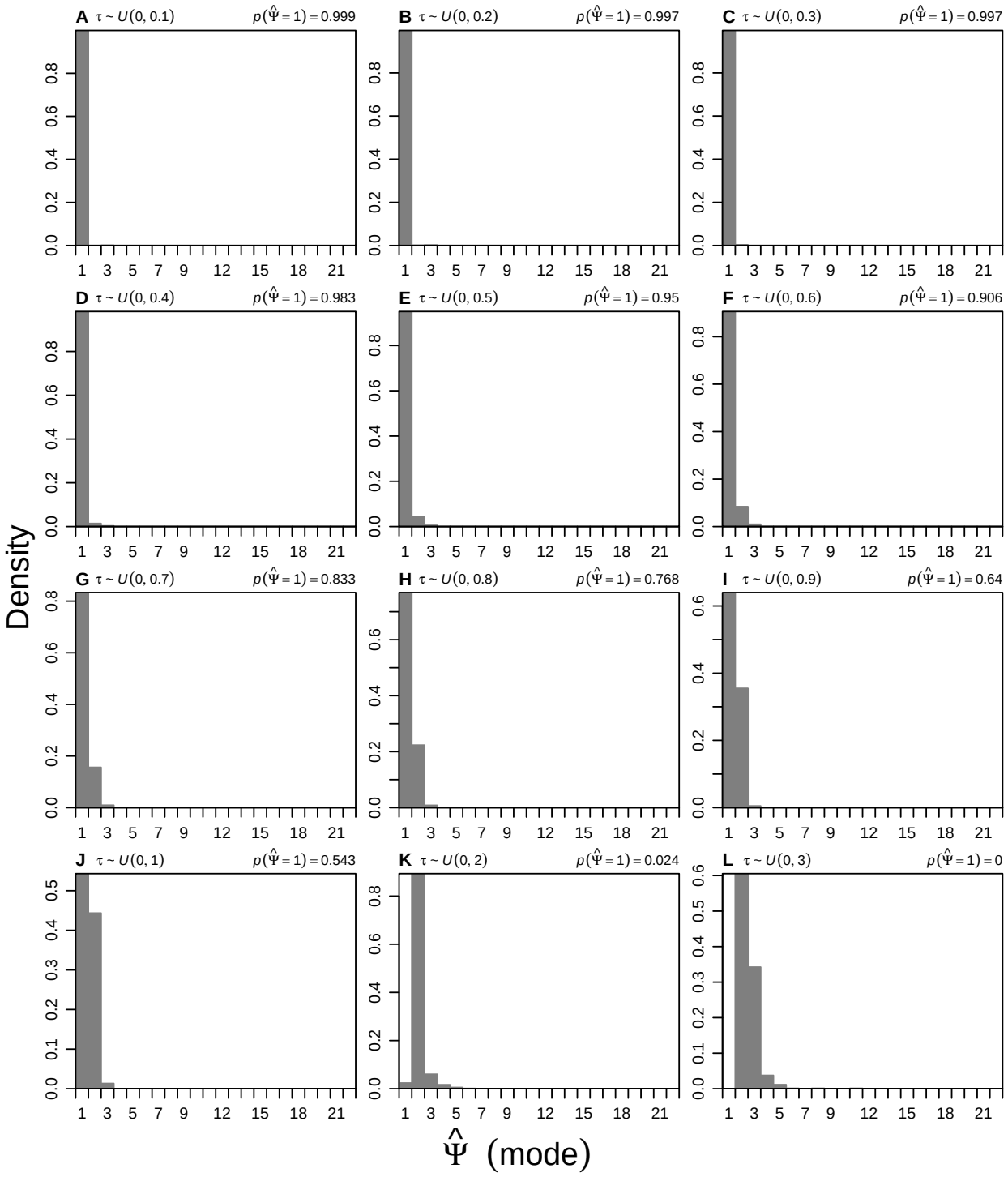

Figure S5.23. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{PLS}$. Note, the logistic regression adjustment provided with `msBayes` failed for approximately 2-30% of the simulations. The failure rate was positively correlated with the number of posterior samples with $\Psi = 1$. Thus, these plots are likely biased towards larger Ψ. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

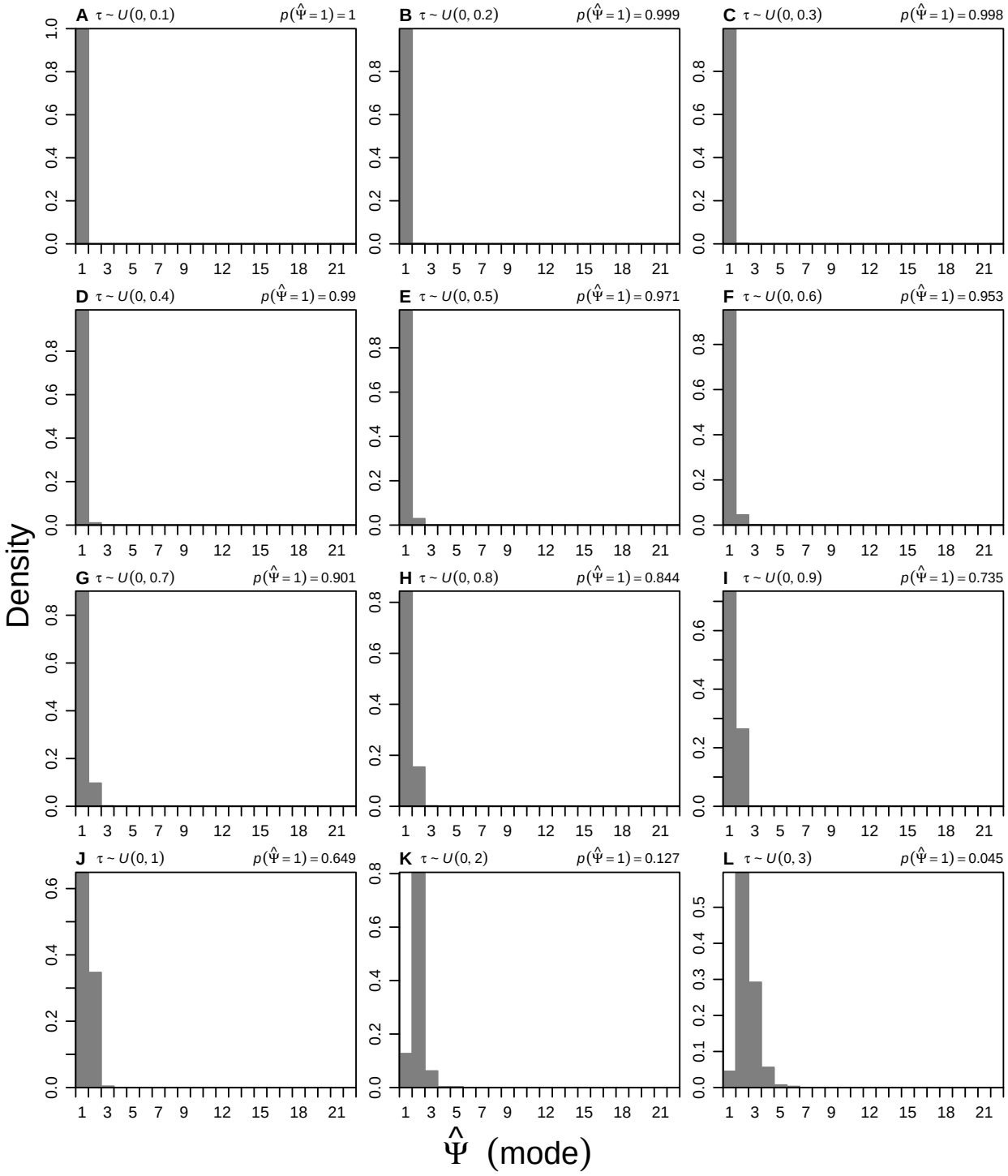

Figure S5.24. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

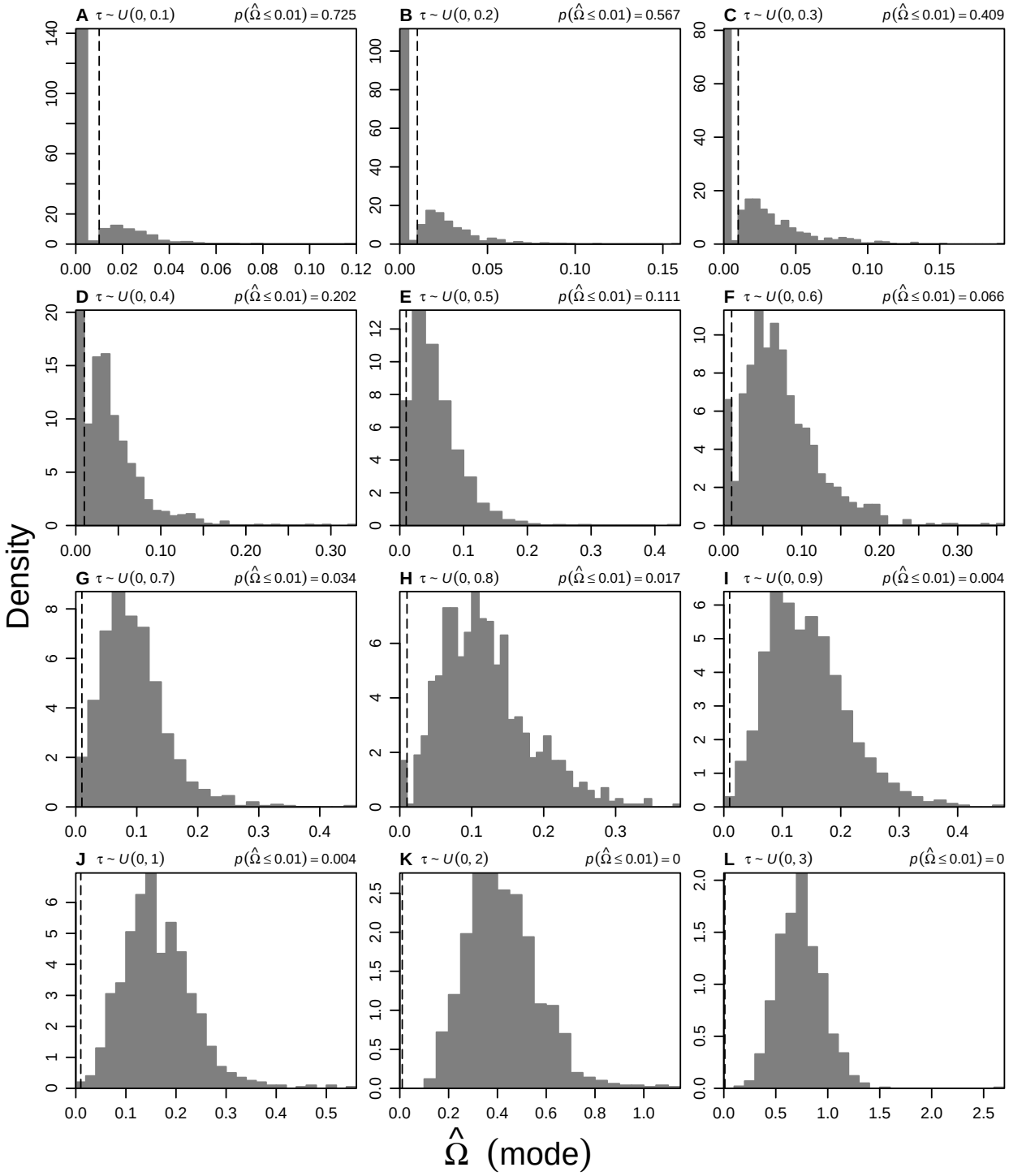

Figure S5.25. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

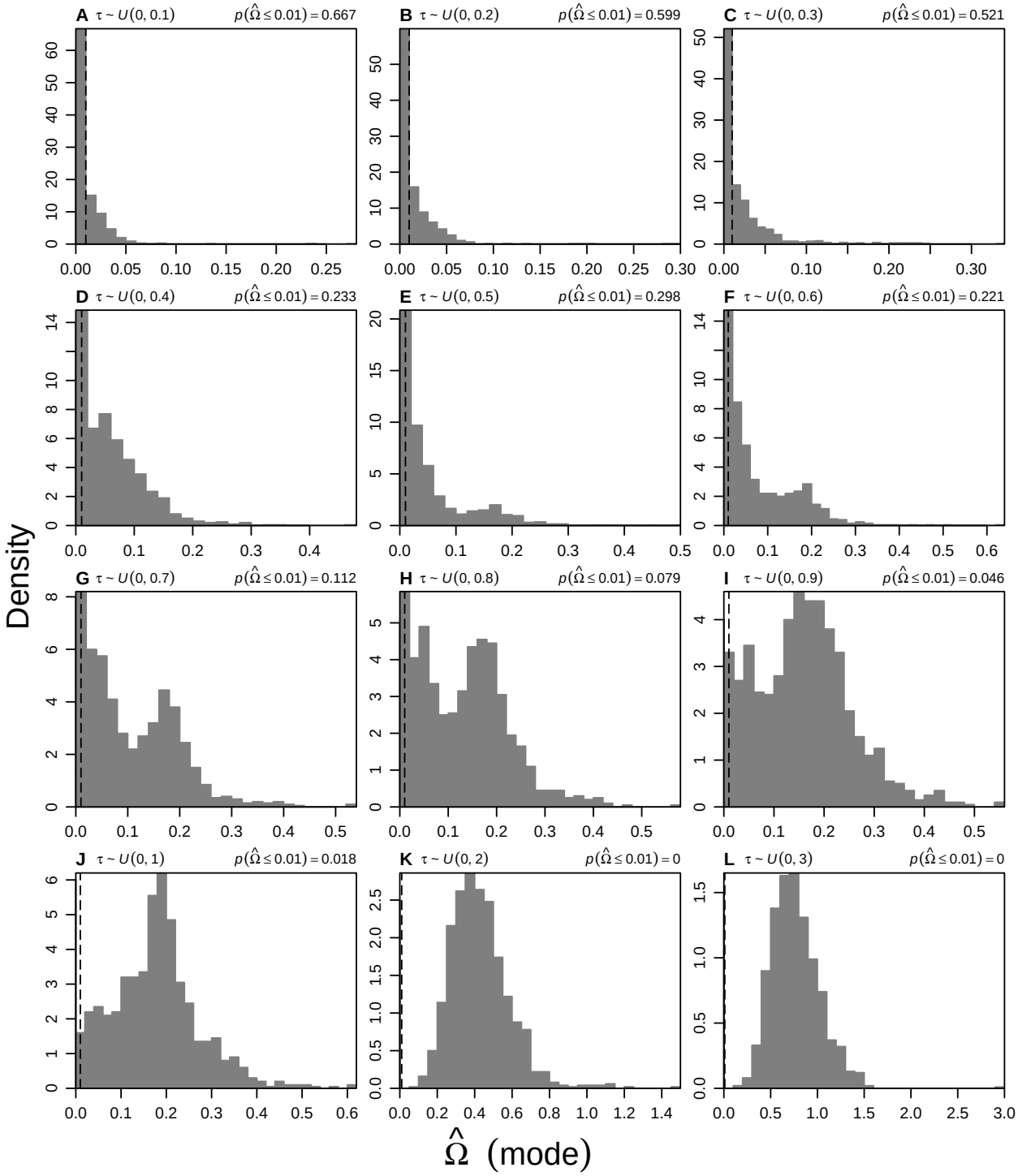

Figure S5.26. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

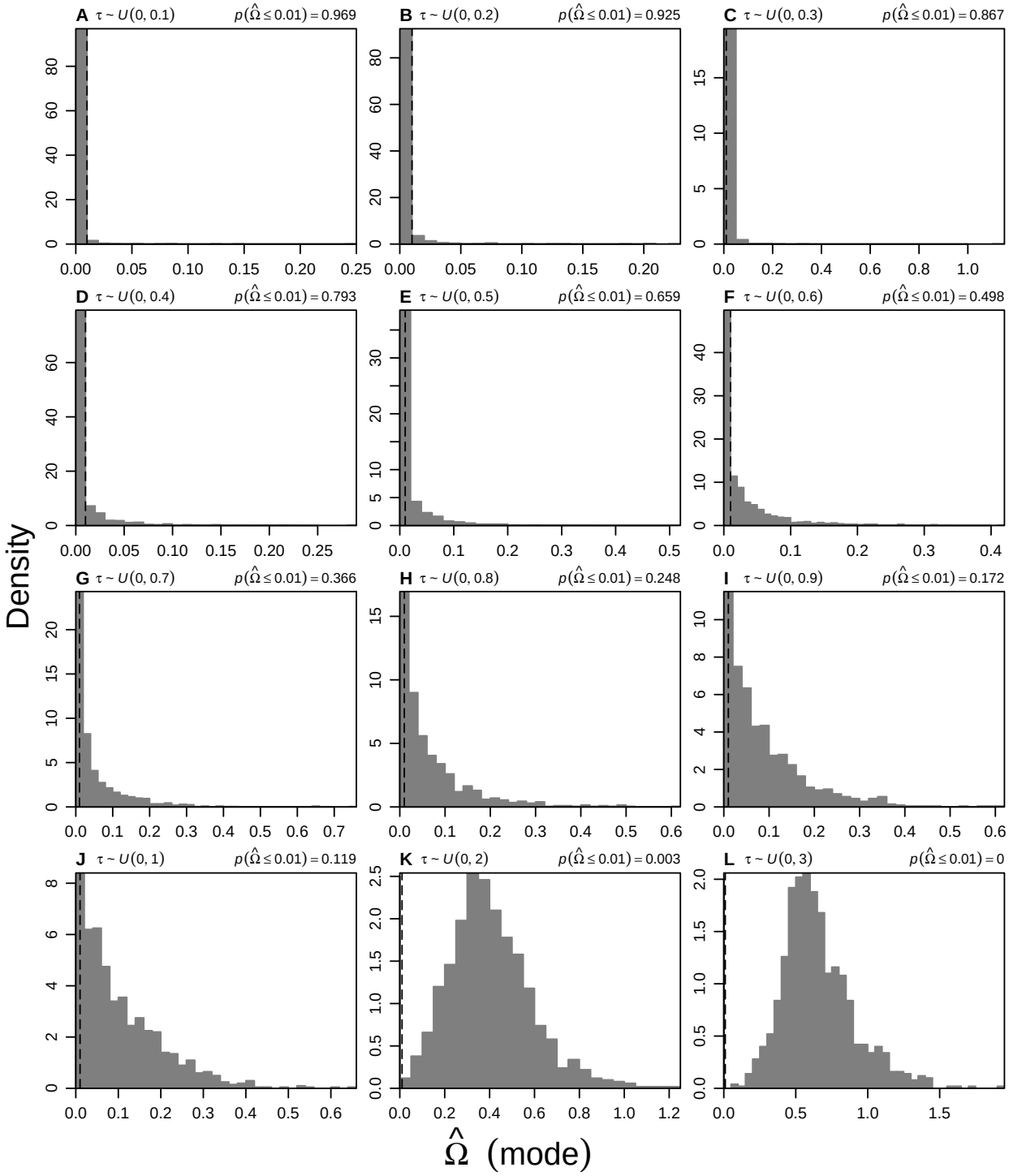

Figure S5.27. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

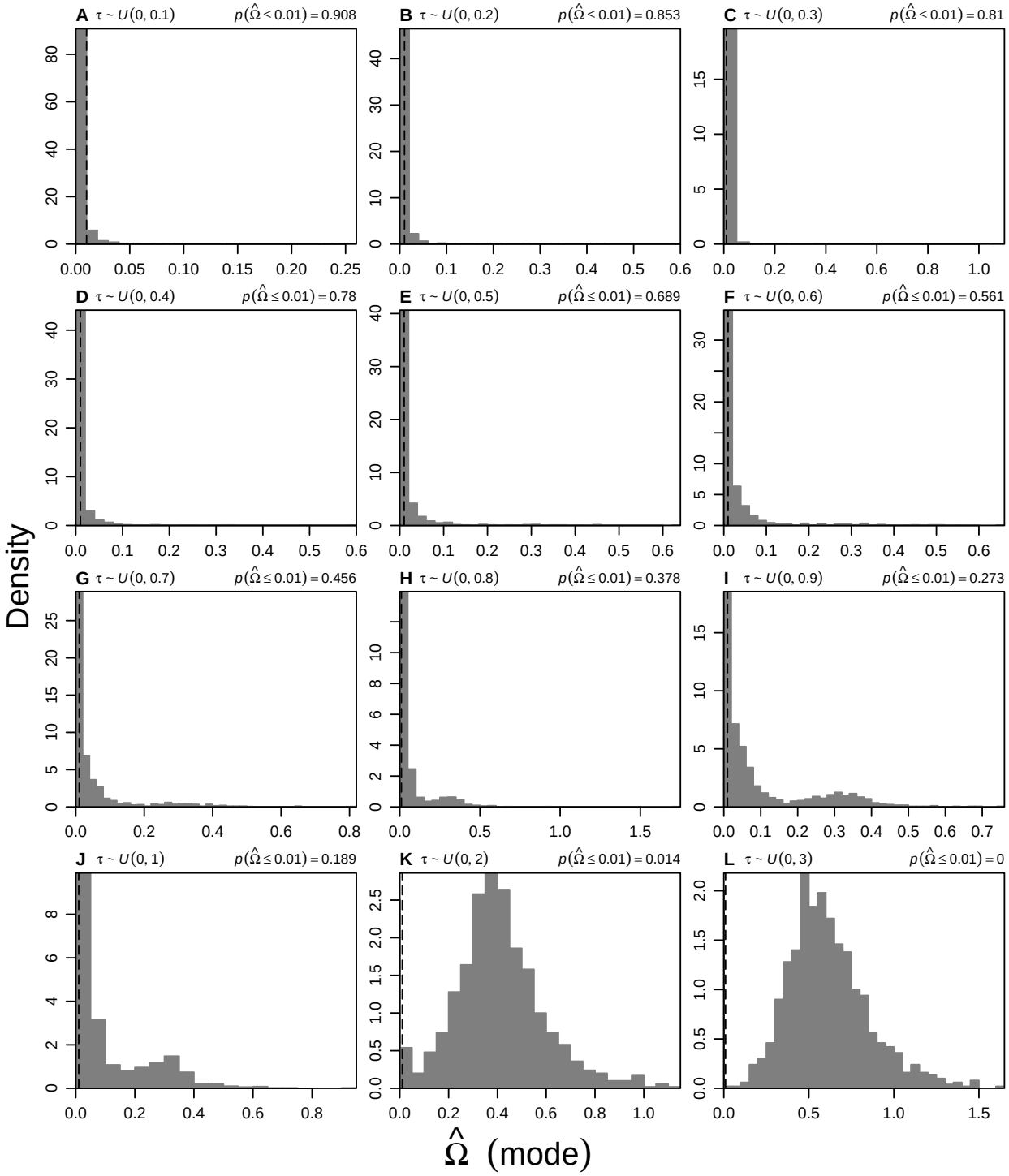

Figure S5.28. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

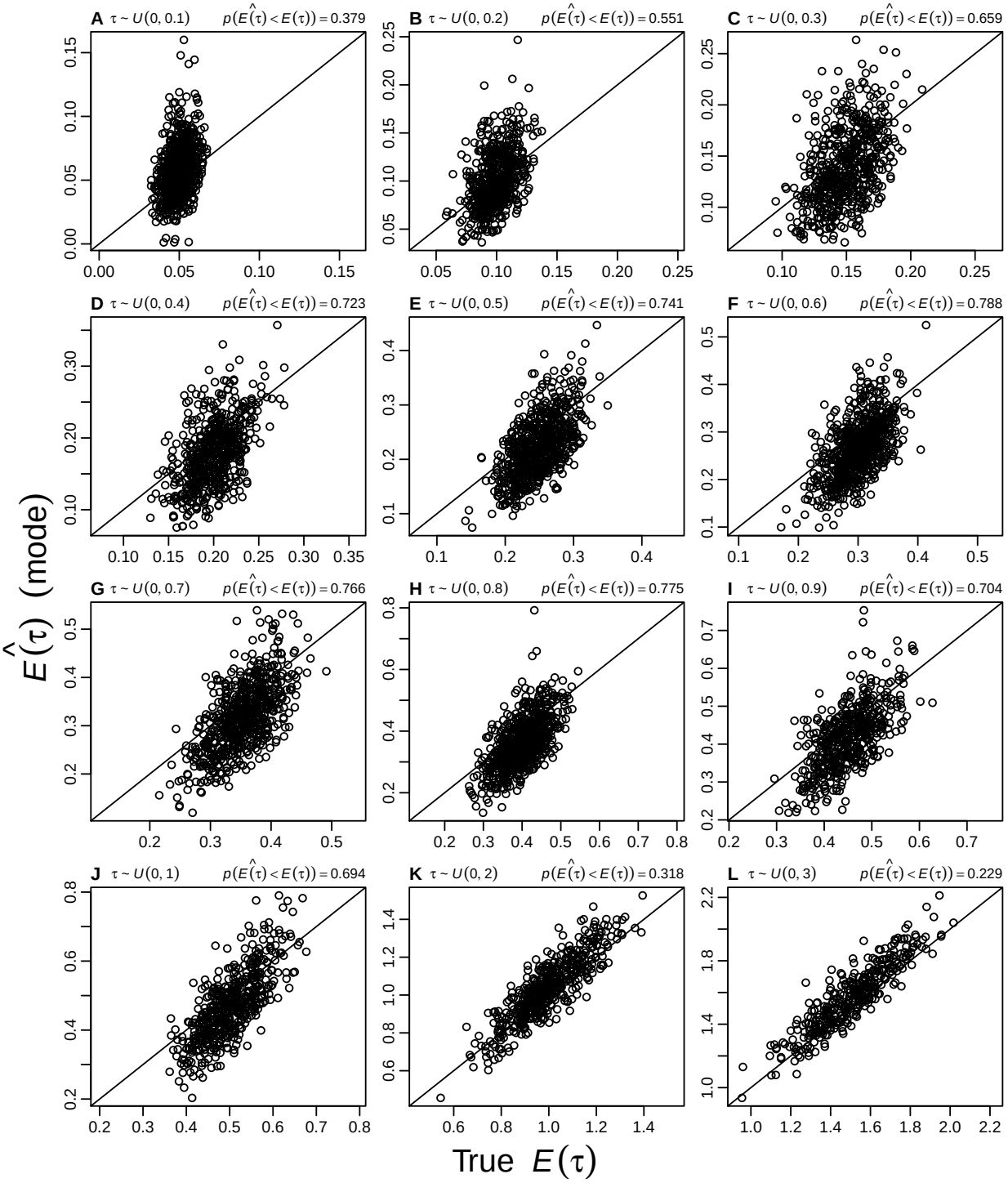

Figure S5.29. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

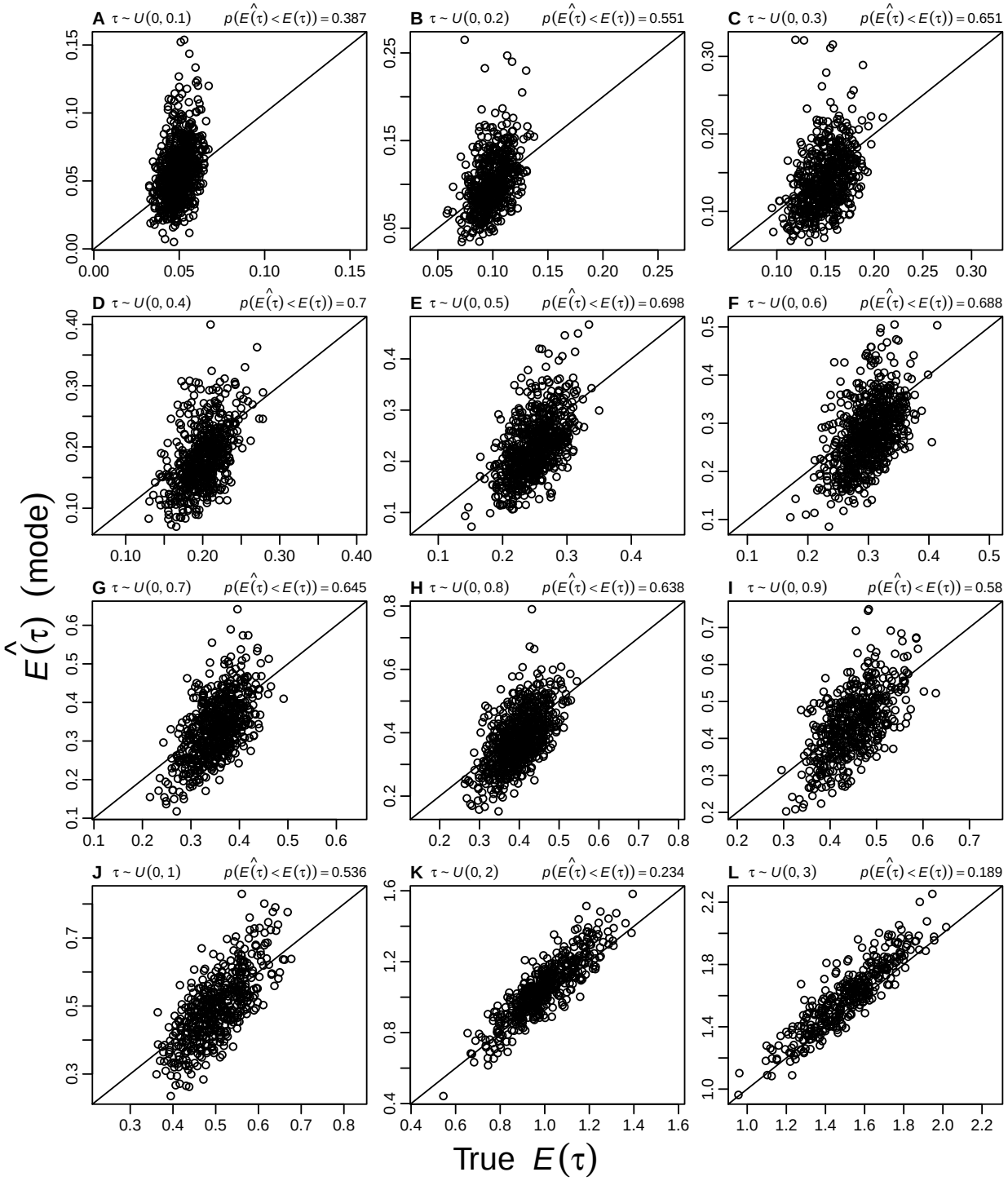

Figure S5.30. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

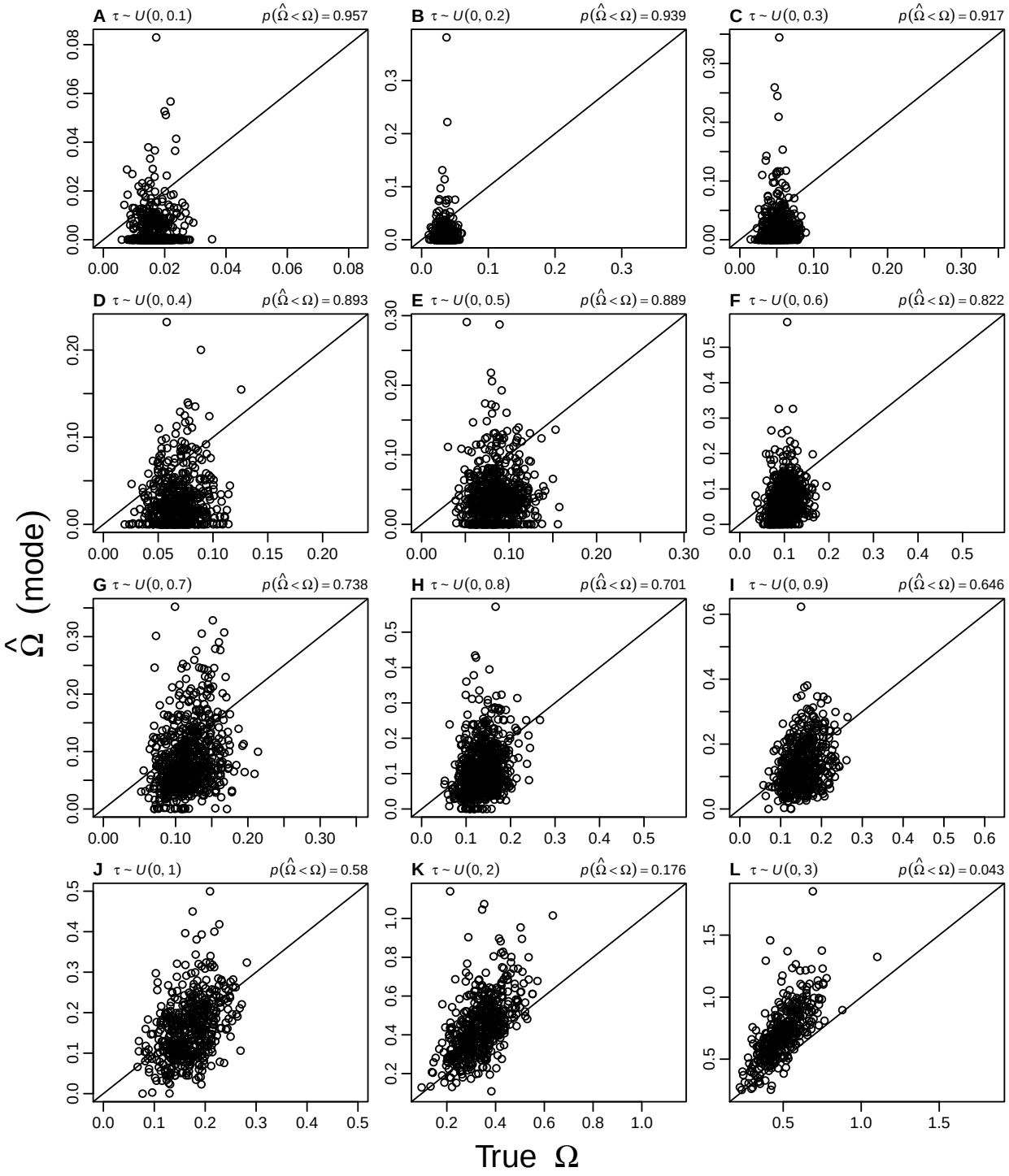

Figure S5.31. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

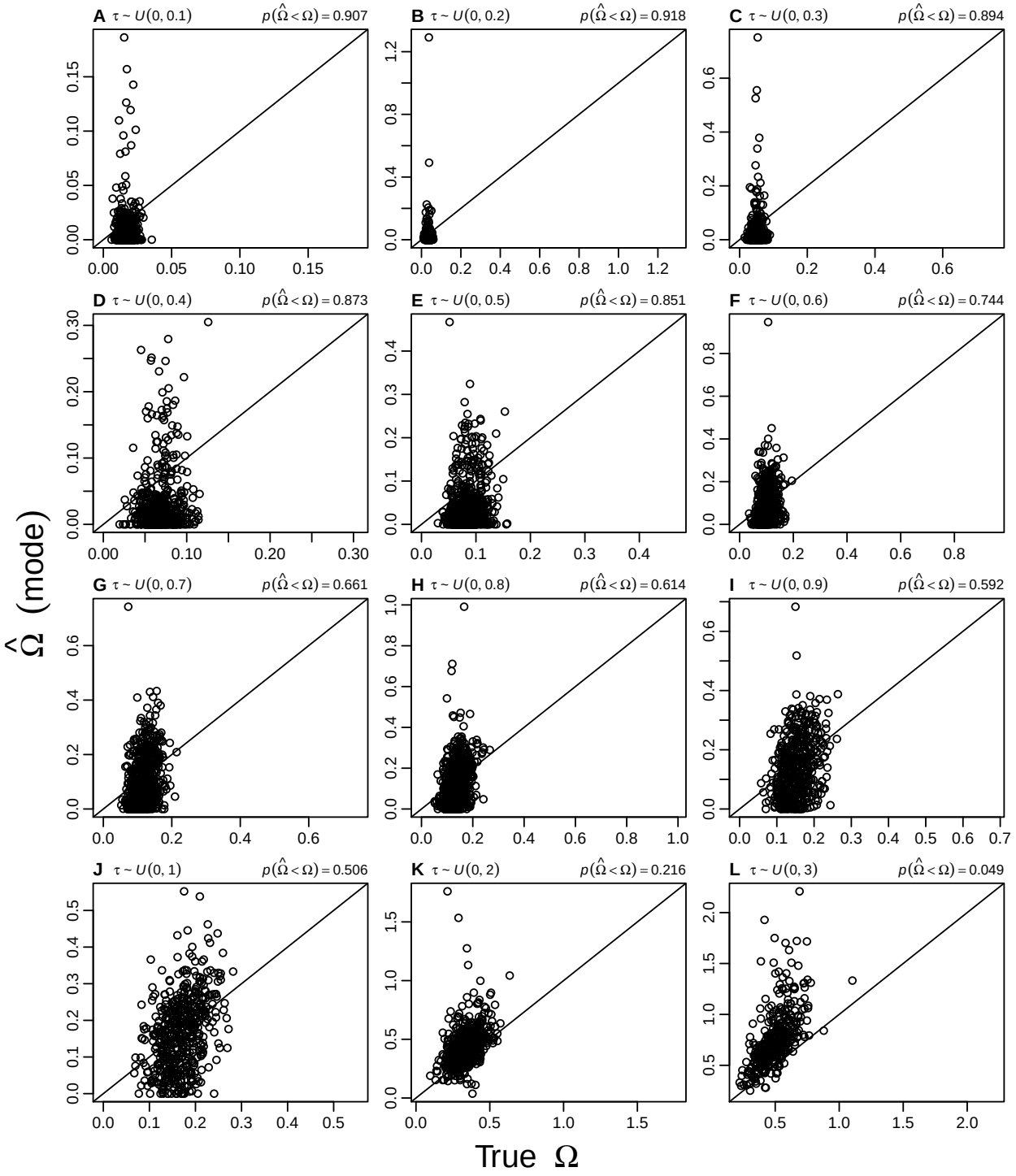

Figure S5.32. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

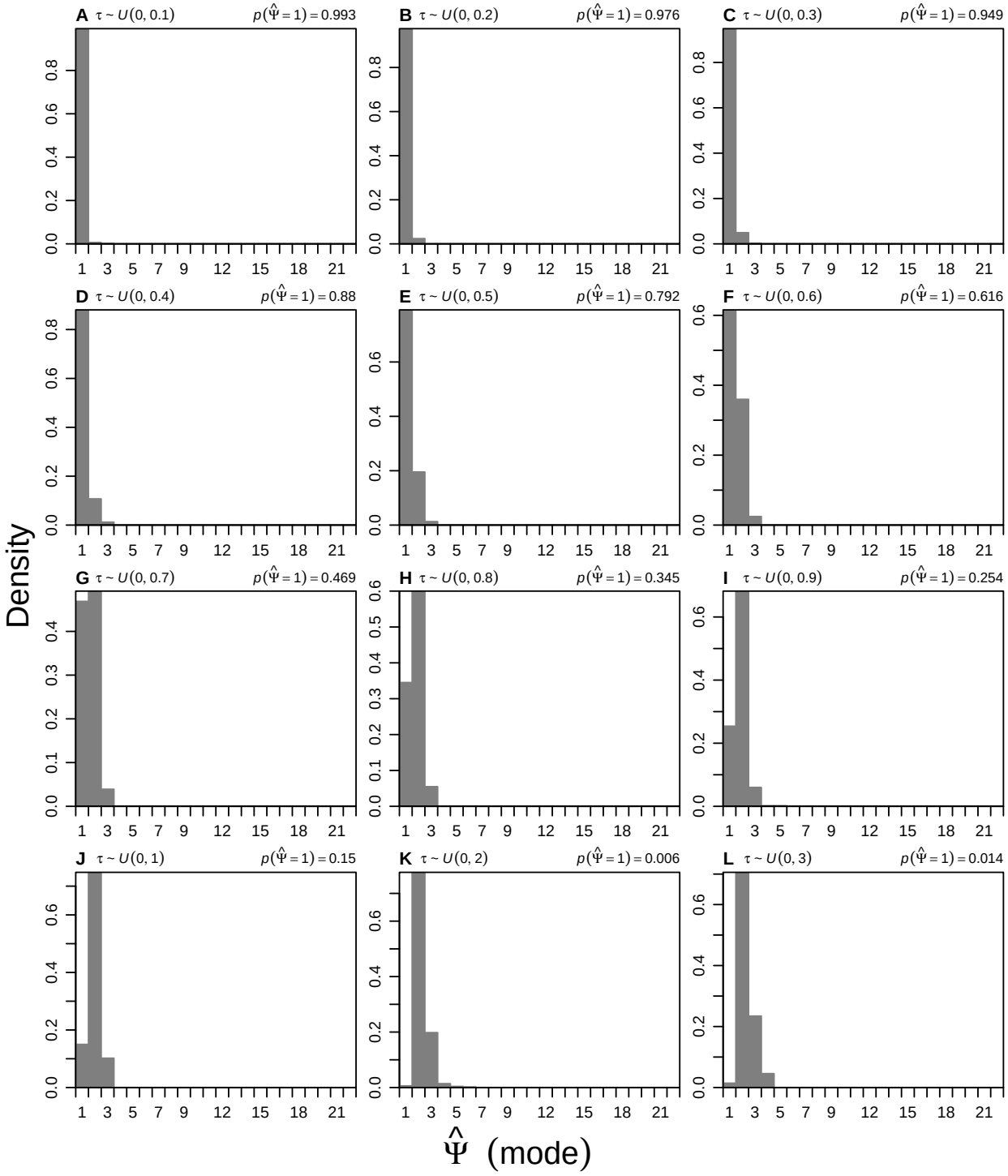

Figure S5.33. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

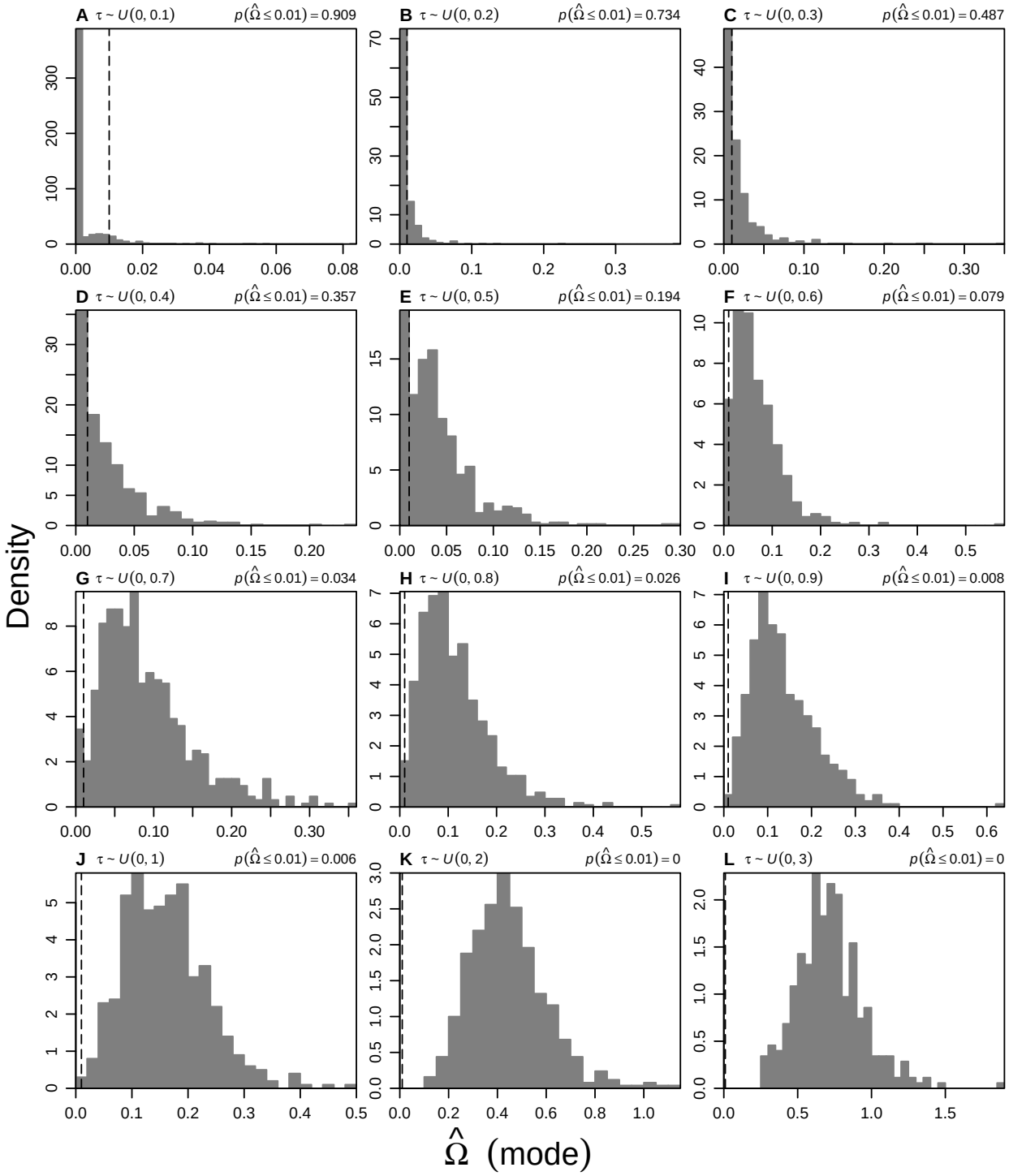

Figure S5.34. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

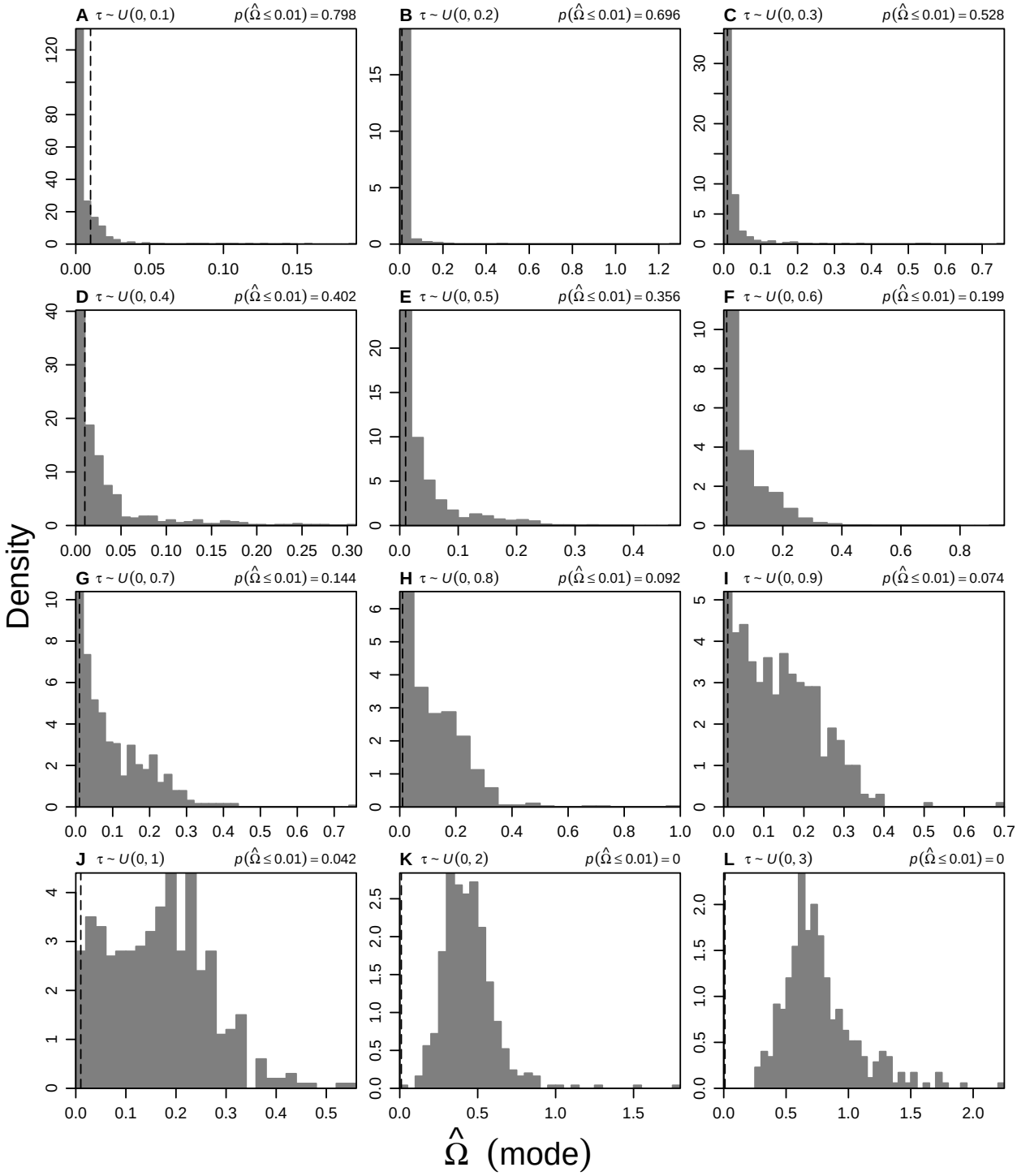

Figure S5.35. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

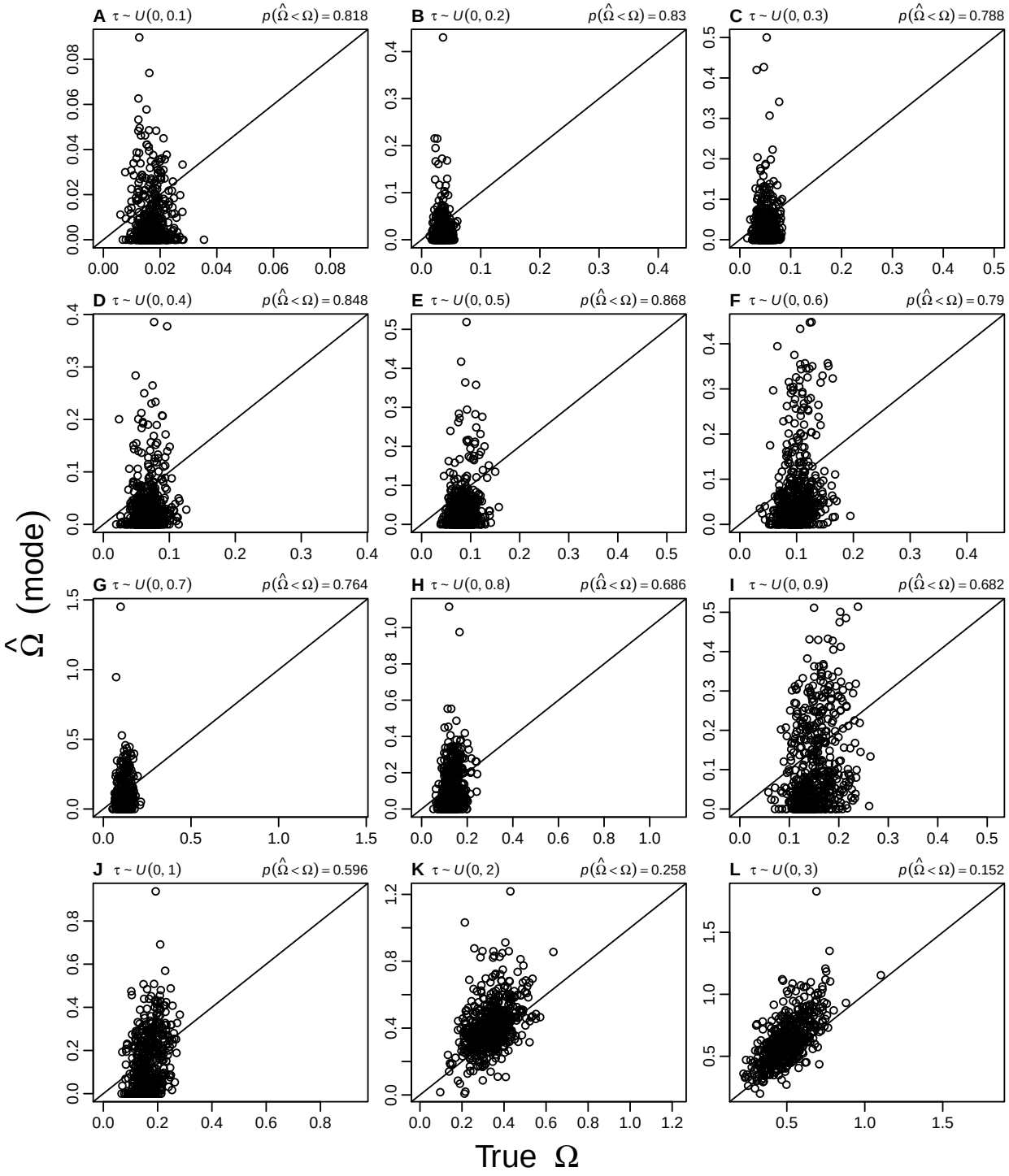

Figure S5.36. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.01, 0.02)$.

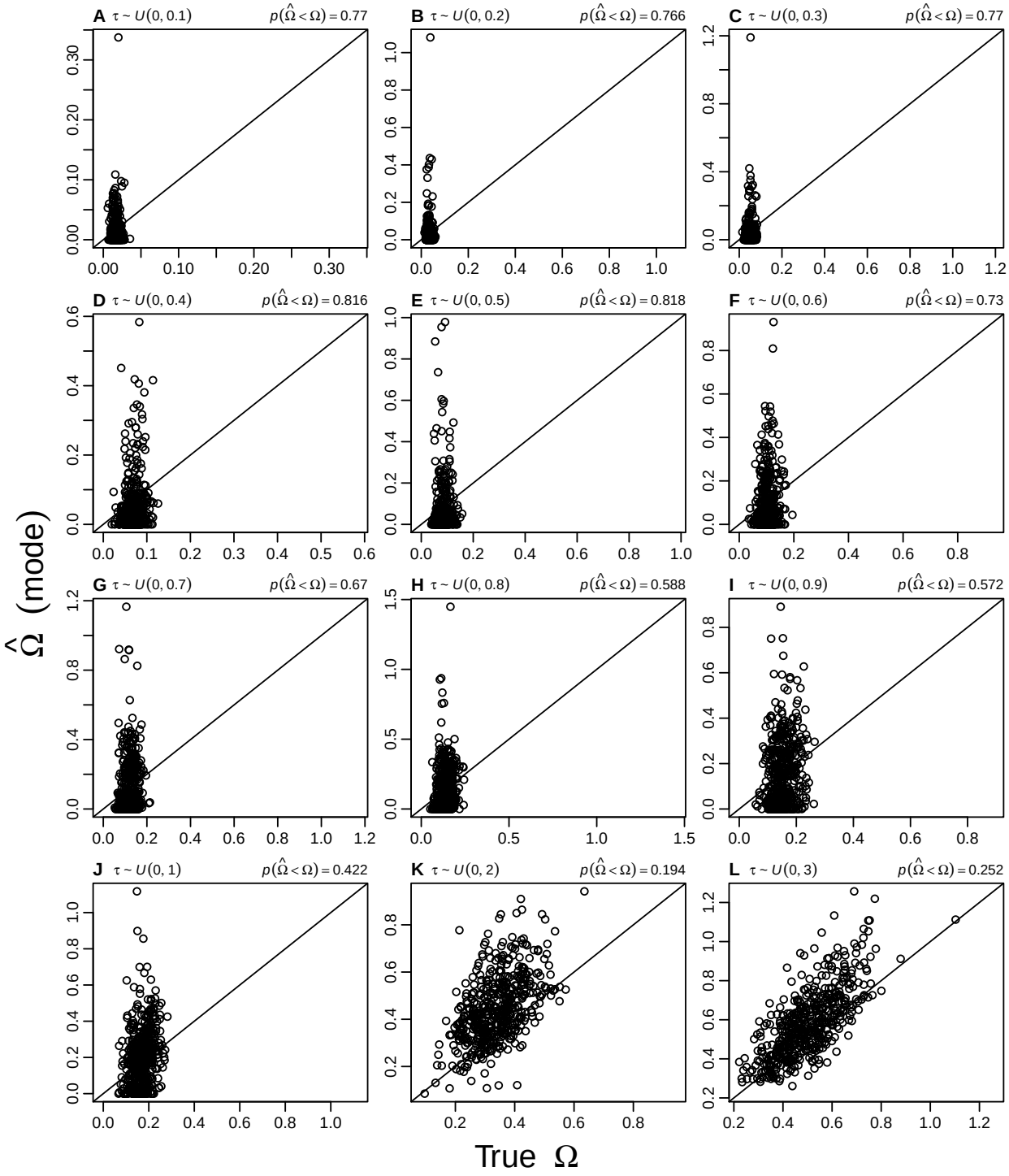

Figure S5.37. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.01, 0.02)$.

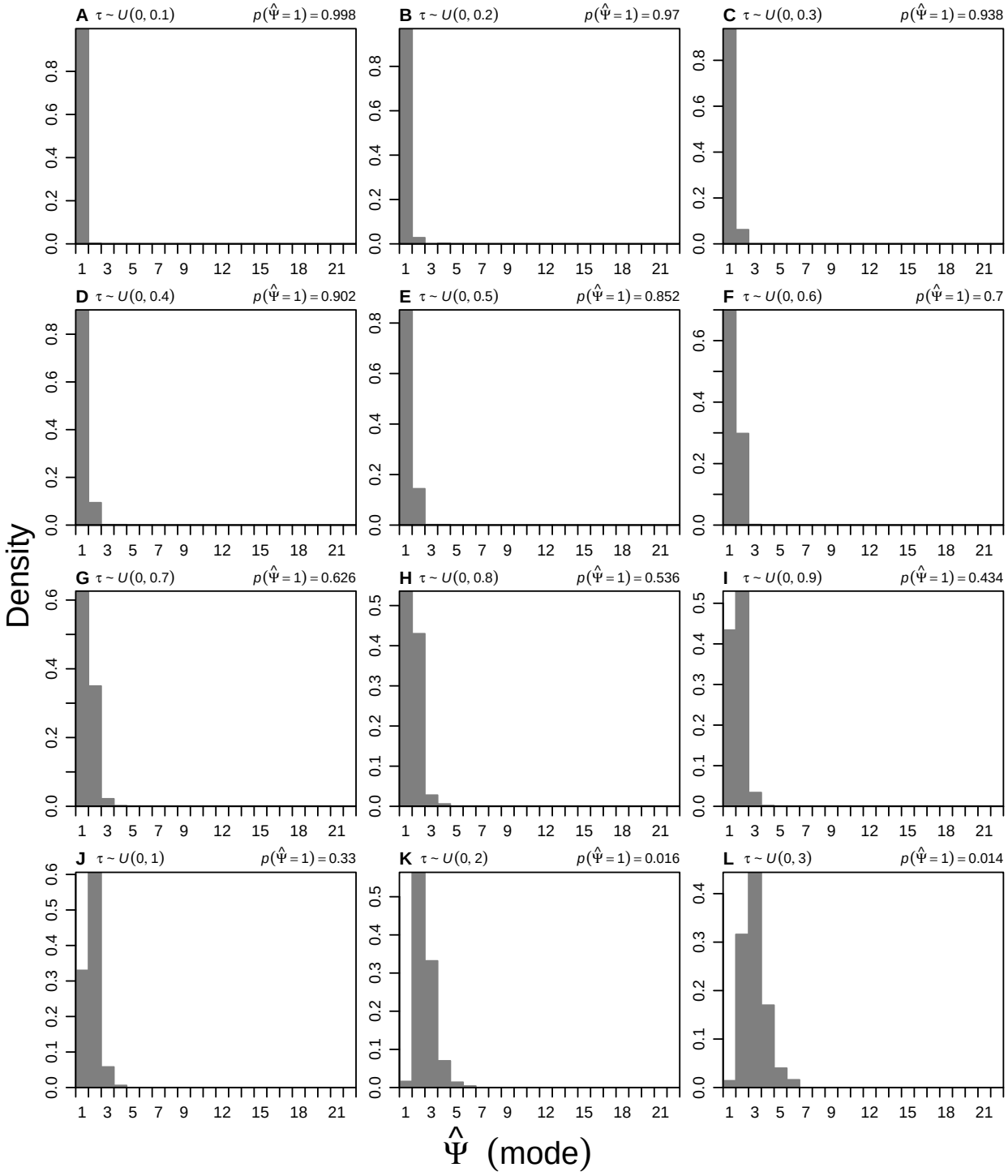

Figure S5.38. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.01, 0.02)$.

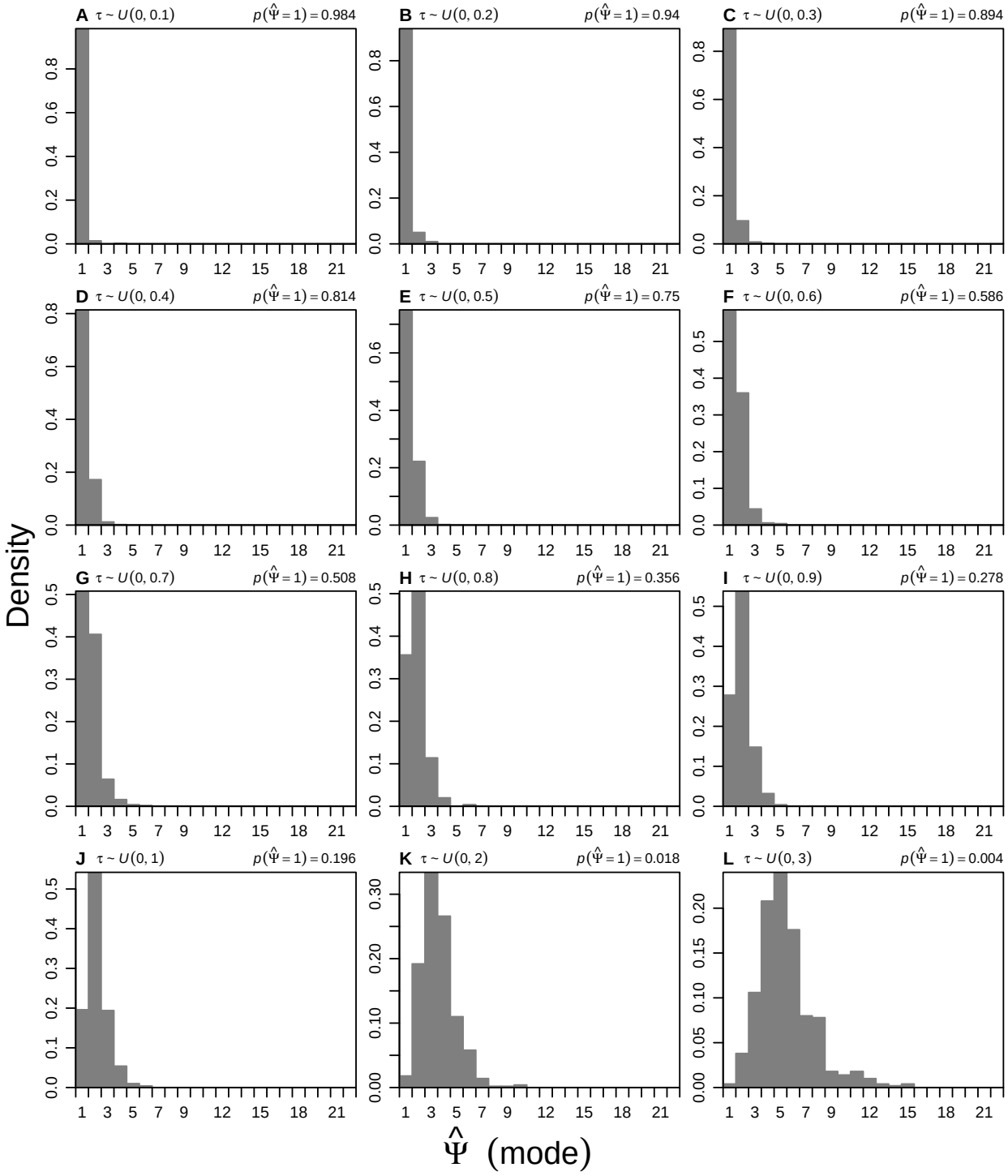

Figure S5.39. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.01, 0.02)$.

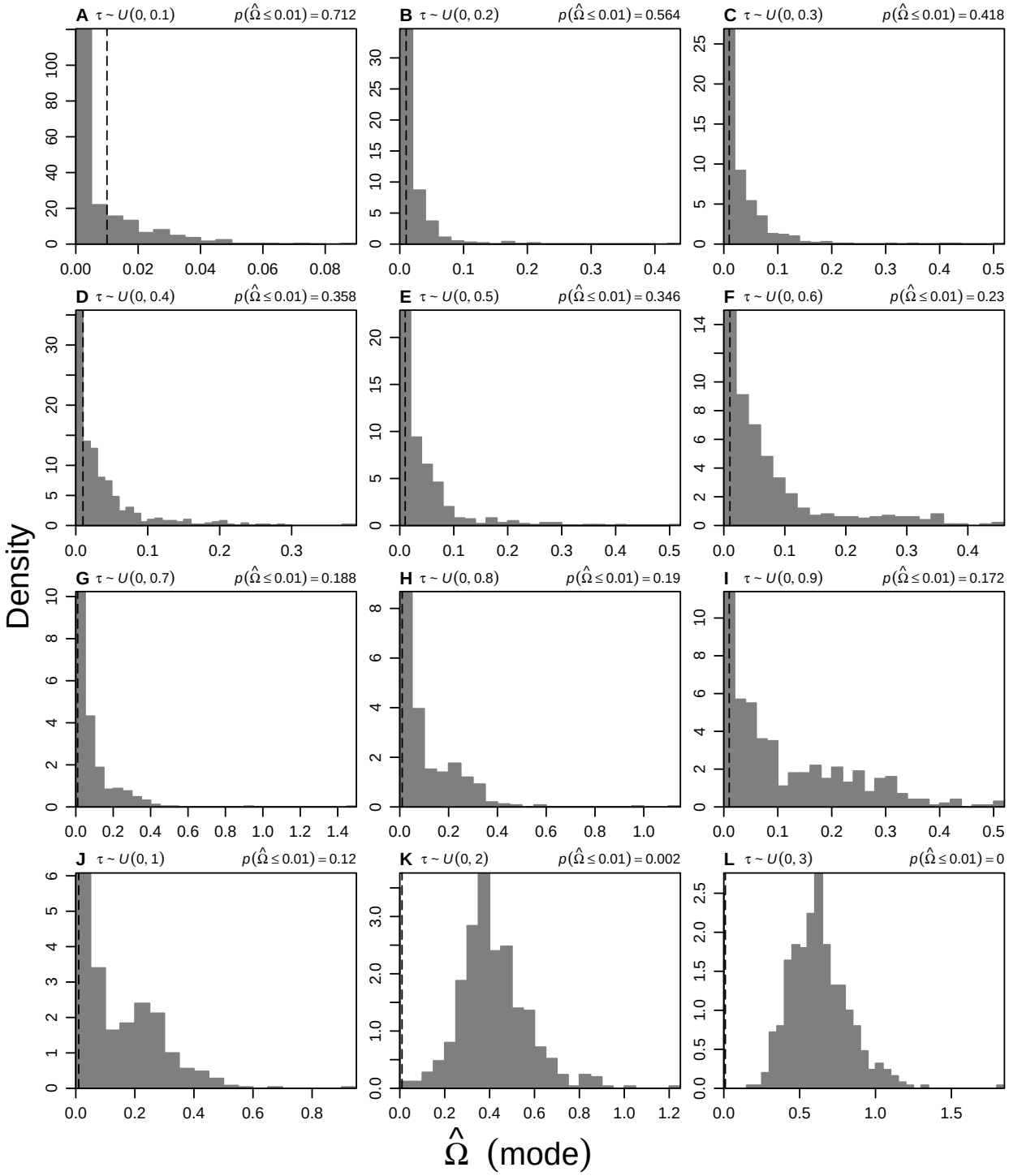

Figure S5.40. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.01, 0.02)$.

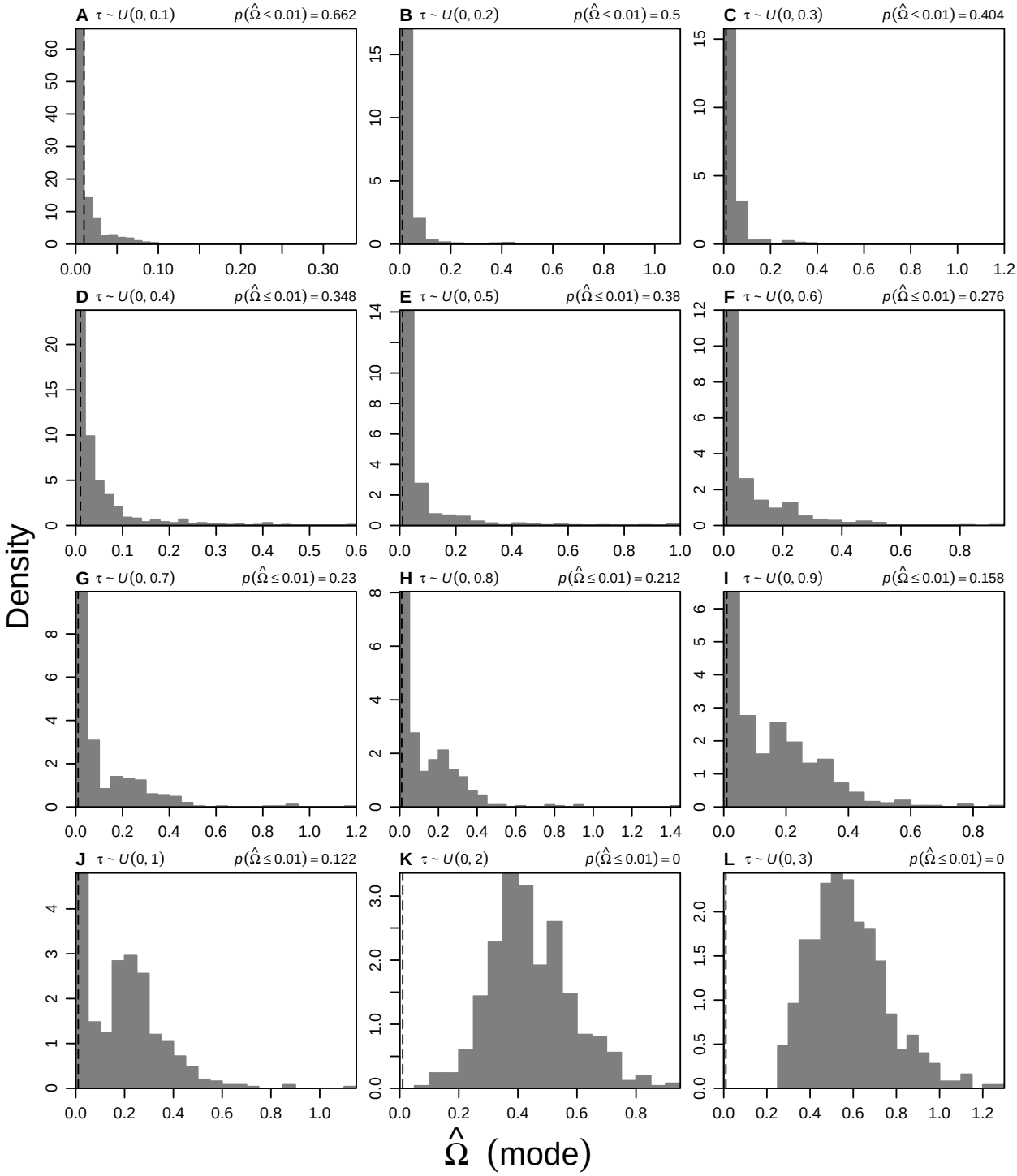

Figure S5.41. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.01, 0.02)$.

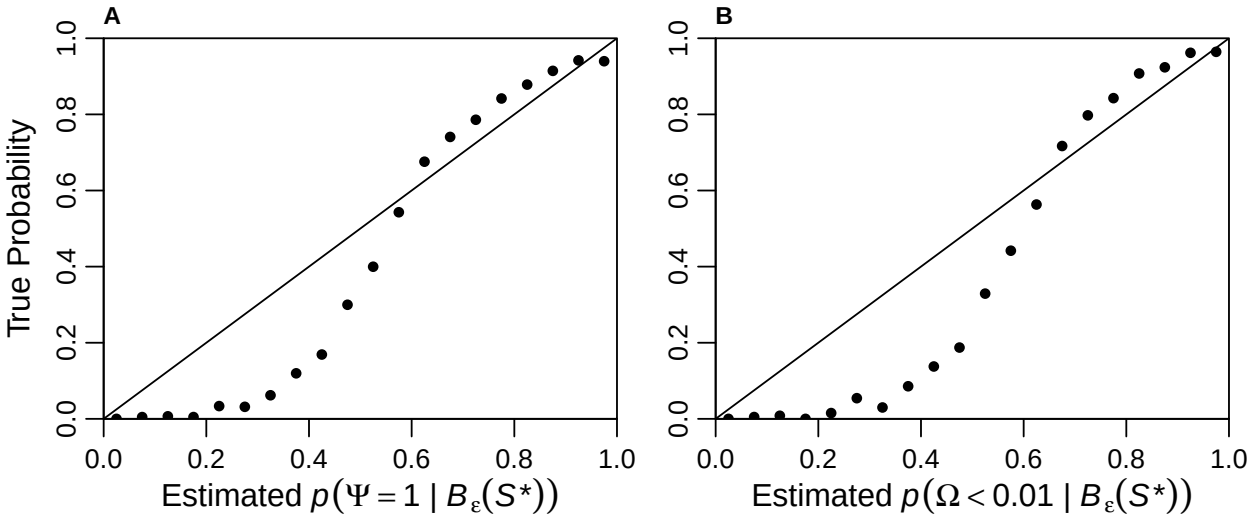

Figure S5.42. The relationship between the posterior and true probability of (A) $\Psi = 1$ and (B) $\Omega < 0.01$ based on 100,000 simulations. The results are based on the unadjusted posterior sample ($\mathcal{P}_\epsilon$) from each simulation replicate. All simulated replicates were generated under the model prior (i.e., the ideal situation where the prior model is correct). Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0001, 0.05)$, and $\theta_A \sim U(0.01, 0.025)$, and the number of samples from the prior was 2×10^6. The simulated data structure was 10 population pairs, with a single 1000 bp locus sampled from 10 individuals from each population. The 100,000 estimates of the posterior probability of one divergence event were assigned to 20 bins of width 0.05. The estimated $p(\Psi = 1|B_\epsilon(\mathbf{S}^*))$ of each bin is plotted against the proportion of replicates in that bin with a true value of $\Psi = 1$.